

What science for Europe's 1992?

An acid test of the European Community's intentions in support of science after 1992 is its willingness to respond to an imaginative proposal from radioastronomers.

EVERYBODY in Europe is being conditioned by the European Commission to believe that 1992 will mark the arrival of the authentic common market in which goods and services will be traded freely, in which capital and people will flow with hardly any artificial interference to the places at which they will be most productive and in which prosperity will grow by leaps and bounds. It is a brave prospect, if somewhat tarnished by the knowledge that this is what should have happened more than 30 years ago, when the European Community was first formed, and by the suspicion that member governments will not instantly forget the stratagems they have learned for replacing tariff barriers by less tangible restraints of trade. Meanwhile, very little has been done to pose intelligently, let alone to answer, the question of what 1992 will do for science and technology. Will 1992 come and go without the question being raised?

With a budget exceeding \$1,000 million a year, the European Commission (the Community's executive branch) is already a big player in science and technology, but most of its present activities in the field will be inappropriate in the circumstances it hopes to have created by 1992. Most of the Commission's energy and money goes into industrially oriented projects such as ESPRIT, the information technology enterprise, in which grants for research and development are made to consortia of research and development groups, mostly industrial but with some academic participation. The goals of these projects are supposed to fall short of the development of products that can be sold, but success is meant to put their participants within reach of profitable markets. As things are, research and development projects qualify for the Commission's support only if the groups they involve are drawn from more than one member state. The rationale for this requirement is that of assisting technical change in a manner that spreads the benefits through more than one member state. The same condition must be satisfied by those in basic research who seek support from the Commission, which is spending upwards of \$150 million a year on projects spanning intra-Community frontiers. This pattern of spending, logical enough in present circumstances, will make no sense if 1992 does indeed accomplish what the Commission hopes.

Reasons

The reasons are straightforward. If 1992 means anything at all, it means that Europe is reconciled to the most efficient division of labour between European manufacturers (of software and hardware alike). Gone will be the time when it could be a legitimate policy objective of either member governments or the commission to ensure that, say, the electronics industry continues to flourish in this country or that. Ideally, this would imply that the Commission should be given powers to ensure that its member governments do not, by subterfuge, use national funds to subsidize the technical development of their own manufacturers. (The use of defence research and development contracts for this purpose is the most serious loophole, which should be closed.)

But, by the same test, the Commission itself should abandon its present pattern of support for industrial consortia spanning Europe's internal frontiers. If the intention is that industrial

companies should compete freely with each other, it will make no sense that the Commission should be spending considerable sums of money to help weaker companies compete with the more successful.

That will bring comfort to some member governments, the British for example, which have been sceptical of the Commission's plans even in present circumstances. But none of that implies that the Commission can or should pull out of research support altogether. While 1992 will rob the Commission's programmes of applied research of their justification, the case for supporting basic research will be even stronger than it is now. Then, in heady 1992, the Commission will inevitably be the source of support in the last resort for the infrastructure of European science. Its function then should be to seek out and to support projects promising to add to Europe's strength in basic science which, for one reason or another, are not supported by national governments. With 1992 almost upon us, the Commission should be starting now to work in that direction.

Disappointment

That is why the commission's failure to support European radioastronomers' plans for strengthening their VLBI network are a cruel disappointment (see *Nature* 337, 301; 26 January 1989). By general consent, the project is an excellent way of making progress in an important field of basic science, but one which also has a potentially important industrial spin-off. The need is for a European centre at which signals recorded by several European radiotelescopes being used in concert as an interferometer can be electronically compared. Technically, the task is complicated because the amount of data gathered by each telescope can be huge and because, in Europe, there may be more than half a dozen independent signals to be compared. The spin-off is the opportunity to develop electronic machinery that is more suitable for this and other information-processing tasks than that now available (from the United States) and also cheaper. Yet the Commission has turned down the proposal on the grounds that none of its existing schemes for research support can accommodate the radioastronomers' ambitions. The amount of money required to launch the project would be too large a proportion of the budget of the committee called CODEST, which is the chief source of funds for basic research, while there would also be running costs which the Commission has no means of providing.

That puts the bureaucratic cart before the horse. If the Commission has no present scheme for supporting a project whose virtues are widely acknowledged, and which could play an important part in strengthening both European science and the ideal of European collaboration, there is a simple answer: the Commission should devise one. And it should do that quickly, however much it may fear that the VLBI consortium's plans, which entail the establishment of a permanent institution, will set an awkward precedent. For the truth is that institution-building of that kind is crucial to the kind of function that will fall on the Commission's shoulders after 1992. The sooner the Commission faces up to that prospect, the sooner it will persuade people that 1992 is really on the way. □



Research by numbers

British grant-makers unwisely conspire with the British government to tell researchers what to do.

BRITISH basic research should be in better shape in the coming financial year than it has been for a long time. That, at least, is the simple expectation. Since the present prime minister promised in 1980 that the science budget would be "protected", the government has dutifully indexed the annual budget for inflation — but has then required that money should be spent in unproductive ways, usually getting rid of people. And while, on two or three occasions, ministers have decreed that there should be a little extra, the cause has usually been an impending crisis. But in the coming year, there will be more than £100 million extra within the research council system, which will have £826 million at its disposal (see *Nature* 337, 587; 1989). So will the year ahead see the beginnings of a resurgence?

Not necessarily. Money is only a necessary, not always a sufficient, guarantor of the volume and quality of research. The privations of the past decade will not quickly lose their effect, while the still-continuing reorganization of British universities and other establishments means continued distraction. Even the extra money is not quite as extra as it seems. Of the total of £107 million, only \$10 million extra will be spent on research grants to people who, fired by a bright idea not advertised in advance, apply for them successfully. Apart from the dismal £14 million for "restructuring", the euphemism for paying people off, the rest of the extra funds will be spent in ways determined in advance either by the research councils (which have taken a sudden liking to what are called interdisciplinary research centres) or by the government (whose encouragement of Antarctic research, for example, first apparent after the Falklands War, has grown marvellously since ozone depletion was first measured from the ground at the British base at Halley Bay). Too much of the extra money will go on predetermined projects.

In *dirigiste* contemporary Britain, such a pattern of events may not be remarkable, which is why it seemed a sign of grace last week that the British government seemed for a time unhappy that individual researchers would win such meagre benefits from its generosity. In the event, the government, as governments will, denied even a scintilla of internal disagreement — but then acknowledged that it exists by announcing a review of the research councils' policy of putting substantial sums of new money into interdisciplinary research centres.

That is a strange development. The fondness for spending research money on autonomous research centres at universities stems from the example of successful investments in the United States and from more numerous but more modest innovations by the Wolfson Foundation in Britain. But in typical British fashion, people are looking to the new research centres not merely for research, but for the solution of related industrial problems and even the training of industrial researchers in new techniques. The themes around which the centres are organized are determined by the research councils. With the first centre (in superconductivity) established only last year, nobody can yet know how successful will be their management, in which related industries are meant to have an important voice. Nothing much has been decided about their long-term future.

If Mrs Margaret Thatcher, or some other minister, has reservations about the wisdom of the past year's rash of decisions to create more of these centres, that is entirely creditable. The research councils, supposedly the custodians of academically based research, have exacerbated the recent crisis by habitually putting too much of their resources into their own activities. The new research centres, which offer grant managers the convenience of writing one large cheque rather than several small ones, could insidiously perpetuate and even aggravate this error. It is proper that the research councils should be required to defend themselves before going too far down the road. But two ironies

remain. First, the new research centres appear to have been devised so as to anticipate the government's wishes, in particular by seeking to tie academic and industrial research together. And second, if the research councils' fault is that they spend too much of their money on predesignated projects, the government's fondness for earmarking research money is even more conspicuous. □

Supporting universities

What is the obligation of the US federal government to university research?

FEW documents produced by the US government have generated more intense passions in the university research community than the Office of Management and Budget's Circular A-21, which spells out the principles and procedures by which universities can recover the indirect costs associated with research. For every research grant awarded to a university researcher, an extra burden falls on the university's shared facilities: the physical plant, the library, the time of administrators and all the other entities that keep an organization going. Because the federal government pays nothing towards the costs of operating universities, it is accepted that it should meet some of the extra costs arising from research. But how much?

It may be easy to detail the specific costs of researchers' salaries, supplies and equipment used in a research project, but reaching agreement on an appropriate allocation of indirect costs has become increasingly contentious. Universities believe they are being asked to shoulder too large a share of the cost of research, researchers believe their grants are being eroded by the seemingly endlessly rising fraction withheld for indirect costs and the federal government, anxious to reduce costs where possible, is zealously seeking reductions in the softer categories such as "departmental administration" or "student service".

The arguments over Circular A-21 have in some cases degenerated to bickerings in which accountants argue whether professors can legitimately be expected to keep track of their activities in 15-minute intervals. (Such a figure may be some solace to those seeking to justify costs, but can have little relationship to reality in most research settings.) Some universities are constantly grumbling that others are treated over-generously.

The American Association of Universities (AAU), having decided in February 1987 that the issue threatened to get out of hand, assembled a committee to decide how it might be made less divisive. The result, a draft report now being circulated for comment, is an attempt to clear the air and start a reasoned debate about the nature of indirect costs. Its telling premise is that the determination of indirect costs is not simply an accounting decision, but one that cuts to the heart of the federal research enterprise.

One of the AAU committee's specific suggestions is that, in the determination of indirect costs, the operation, maintenance and depreciation of equipment and facilities should be separated from components such as administrative and library services. The argument is that that will be legitimate pressure to increase the facilities component as existing facilities continue to age, and that it will make for better-tempered negotiations between university and federal accountants if the causes of increased costs are identified, not confused with, for example, administrative costs. Another proposal would set a fixed value for the administrative component of indirect costs that would apply to all universities, thus avoiding the regular and unseemly squabbles which the percentage rates provoke.

Details apart, there is a need to move away from creative accounting in determining the financial relationship between universities and the federal government, and towards a more explicit sense of partnership that acknowledges the needs of all parties to the research effort. The AAU report will at least help to raise the profile and tenor of the debate. □

Arctic chemistry may cause significant ozone loss

- International team returns from Norway
- Duration of polar vortex a key factor

Washington & London

CONDITIONS in the Arctic are ideal for significant depletion of stratospheric ozone. That is the message brought back last week from a six-week mission to Stavanger, Norway, led by the US National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). But the expedition did not remain in the Arctic long enough to observe any stratospheric ozone loss, and the extent of any depletion will depend on how long the polar vortex holds together.

Although the mechanisms for the loss of stratospheric ozone have become increasingly clear in the Antarctic, the picture in the Arctic is more complex. Higher temperatures at the North Pole produce fewer Polar Stratospheric Clouds (PSCs) in which the normally inactive forms of chlorine (HCl and ClONO_2) are transformed into active ClO , which can destroy ozone in the presence of sunlight.

In addition, in contrast with conditions at the South Pole, the Arctic vortex typically breaks up before the sunlight necessary for ozone destruction strikes it. Despite these differences, 5 to 6 per cent decreases of total stratospheric ozone have been seen between 40°N and 60°N during winter, suggesting some disturbance in the Arctic atmosphere.

Beginning in early January, the Airborne Arctic Stratosphere Expedition assembled at Stavanger to make chemical measurements of the atmosphere over the pole. The campaign was coordinated by NASA and cosponsored by NOAA, the National Science Foundation and the US Chemical Manufacturers Association. The UK Meteorological Office, the UK Department of the Environment and the Norwegian Meteorological Institute provided support and researchers, as did universities in the United States, the United Kingdom, Norway and West Germany. Two specially equipped aircraft, an ER-2 capable of flying up to 19 km and a DC-8 flew 14 missions each. More data came from a network of balloon-launched sondes.

After a frustrating period in early January when the vortex was located beyond the range of the ER-2 aircraft (see *Nature* 337, 492; 1988), the expedition found two types of PSC, both apparently capable of producing the active chlorine species. The most common PSCs were those of Type I, which form at -78°C and

are thought to contain nitric acid trihydrate. Occasionally water-ice Type II PSCs that form at -85°C were observed.

By early February, ClO mixing ratios had reached values similar to those seen in the Antarctic, as high as 800 parts per 10^{12} by volume inside the vortex, compared with values closer to 50 parts per 10^{12} outside. BrO radicals were also observed at mixing ratios of 2 to 8 parts per 10^{12} , also similar to levels in the Antarctic vortex.

Stratospheric dehydration and denitrification, believed from Antarctic expeditions to be a necessary condition for the appearance of active chlorine species, was not seen by the Arctic expedition. Although there were localized occurrences of denitrification, these data imply that PSCs may be the crucial factor in the appearance of active chlorine compounds.

In a statement outlining its initial results, the expedition concluded that there was "no unequivocal signature of photochemical loss" of ozone during the mission. But NASA project scientist Robert Watson says that, if the vortex persists, ozone depletion of about 0.5 to 1 per cent a day could occur once the sun strikes it. When that occurs will depend on how long the vortex stays together, and how it moves. According to the UK Meteorological Office, a stratospheric warming that began on 31 January could result in the break-up of the vortex within the next week or two.

Preliminary indications of ozone depletion come from the Canadian Atmosphere and Environment Service (AES), which last week reported a 5 per cent decline in stratospheric ozone as measured from balloon sondes from the Canadian station at Alert. According to Wayne Evans of the Environment Service, PSCs were also observed from Alert. Evans says the Canadian base was not equipped to measure ClO concentrations.

Speaking at a press conference last week, Watson emphasized that the results of the expedition are strong evidence that the atmospheric chemistry in the Arctic, although not as important as in the Antarctic, plays an important and hitherto unacknowledged role in the destruction of global stratospheric ozone. He pointed out that the models used to determine restrictions on chlorofluorocarbons adopted in the Montreal Protocol did not take into account heterogeneous chemistry now shown to be occurring at both poles.

Joseph Palca & Philippa Lloyd

Compensation for haemophiliacs

Tokyo

A PANEL set up by Japan's Ministry of Health and Welfare last week approved the first relief payments to haemophiliacs infected with the AIDS virus, thus ending a long controversy over the government's responsibility for screening blood products.

The National Association of Haemophiliacs has been fighting for compensation for alleged government negligence in failing to ensure a blood supply free from the AIDS virus (see *Nature* 333, 5; 1988). Most of those in Japan infected by the virus are believed to be haemophiliacs and members of their families. Relief funds have been donated by blood product manufacturers and provide monthly allowances of ¥208,900 (\$1,600) for infected people and monthly compensation of ¥156,900 (\$1,225) to families where a main income earner has died from AIDS.

Also last week, Japan's 'AIDS prevention law' took effect after a long spell stalled in the Diet last year. The law contains controversial clauses, attacked as infringing human rights, that require doctors to report details of patients who test positive for the AIDS virus. Names and addresses must be registered if a seropositive person is judged likely to infect other people. Mandatory registrants include drug users and those who regularly have unprotected sex with two or more persons other than their spouses.

Alun Anderson

New Delta rocket off the ground

Washington

THE first of a new family of rockets that will provide the United States with an alternative to the space shuttle blasted off successfully from Cape Canaveral last week. The Delta II rocket, which placed a military navigational satellite in orbit, is one of a new fleet of powerful expendable rockets bought by the Pentagon to reduce their dependence on the shuttle. A Delta II can put a 1.8-tonne payload in geostationary orbit.

Next month should also see the first launch of an upgraded Titan rocket, Titan IV, which will be able to launch a 4.5-tonne payload into a stationary orbit, the same lift capacity as the shuttle.

Last week's payload was the first in the Navstar Global Positioning System, an array of 21 satellites in six orbital planes that will allow US and allied forces on the ground, at sea and in the air to determine position and altitude with great accuracy, at any time, anywhere on the globe. The Navstar system will also be used for civilian navigational purposes. David Swinbanks

West Germany cracks down on exports of weapons gas

Munich

WEST German Chancellor Helmut Kohl withstood a blistering attack for his behaviour in the continuing Libyan chemical weapons scandal as his government introduced new measures on 15 February to tighten West German export laws.

Kohl's centre-right coalition government last week issued a report confirming that Bonn was informed as early as 1980 that West German companies might be participating in the building of a chemical weapons factory at Rabta in Libya. (Kohl's government took office only in 1982.) In 1985, the West German com-

pany about the charges earlier than October 1988, then he was remarkably uninformed, said SPD parliamentarians. And if he did know, they asked, why didn't he do anything?

Kohl was criticized by his own coalition partners in the parliamentary debate, raising the spectre of a breach within the six-year-old coalition. Although foreign and West German observers foresee no immediate change of government, there is some speculation that the coalition might not survive the 1990 parliamentary elections.

The government's plan for improving export controls includes lengthening the

Export restrictions on chemicals with 'weapons' applications

Compound	United Kingdom	United States	West Germany	East Germany	France
1. Thiodiglycol	■	■	■	■	■
2. Phosphorous oxychloride	■	■	■	■	■
3. Dimethyl methyl phosphonate	■	■	■	■	■
4. Methyl phosphonyl difluoride	■	■	■	■	■
5. Methyl phosphonyl dichloride	■	■	■	■	■
6. Dimethyl phosphite	■	■	■	—	□
7. Phosphorous trichloride	■	■	■	■	■
8. Trimethyl phosphite	■	■	■	—	□
9. Thionyl chloride	—	■	□	—	—
10. 3-Hydroxy-1-methylpiperidine	—	■	□	—	—
11. <i>N,N</i> -diisopropyl- β -aminoethyl chloride	—	■	□	—	—
12. <i>N,N</i> -diisopropyl- β -aminoethane thiol	—	■	□	—	—
13. 3-Quinuclidinol	—	■	□	—	—
14. Potassium fluoride	■	■	□	—	—
15. 2-Chloroethanol	■	■	□	■	—
16. Dimethylamine	■	■	□	■	—
17. Dimethylamine hydrochloride	—	■	□	—	—

■: Export licences required. ■: Licence required for export to certain countries. □: Export licences to be required in near future. —: No licence required. Data are courtesy of the Geneva Disarmament Conference. The first five compounds on the list include precursors of mustard gas and the nerve gases tabun and sarin. The list includes fluoridating agents (14) and solvents (15) with a number of industrial and pharmaceutical uses as well as phospho-organic (6–8) that may serve as nerve gas precursors. (Data supplied by Verband der Chemischen Industrie).

panies involved were first mentioned by name in a confidential report from the West German ambassador in Moscow. The list included Imhausen-Chemie GmbH, the company that allegedly designed the factory and provided West German engineers to help with its construction.

The government now says that it could not have acted sooner to stop exports to Libya. A preliminary investigation was not warranted until July 1988, Chancellor Minister Wolfgang Schäuble insisted last week, adding that Kohl was first informed of the charges on 20 October 1988.

The opposition Social Democrats battered Kohl in an unusually acrimonious televised parliamentary debate. Party leader Hans-Jochen Vogel accused Kohl of "shameful" insensitivity for not acting sooner. If Kohl did not know

list of banned chemicals and equipment, improving communications among export authorities in West Germany, increasing the penalties for those involved in chemical or biological arms production, and increasing threefold the personnel responsible for monitoring West Germany's massive exports. The plan was agreed upon in principle by all parties. Legislation is expected to follow later this year.

West German Foreign Minister Hans-Dietrich Genscher (Free Democrat) acknowledged that unilateral action to ban chemical arms exports is insufficient. He called for rapid progress towards an international, verifiable ban on the manufacture of chemical weapons and on stockpiling.

In a late development on 20 February, foreign ministers of all 12 members of the European Community (EC) voted to require export licences for 8 chemical

Soviet research ship is impounded for "illegal entry"

London

THE Soviet research ship *Akademik Mstislav Keldysh* became the focus of a diplomatic incident two weeks ago, when the harbour authorities at Freetown, Sierra Leone, impounded the ship for allegedly making an illegal entry into the harbour.

The nuclear-powered ship, which belongs to the Soviet Academy of Sciences, carries two submarines for sea-bed exploration and a scientific staff of 100. According to the Soviet Foreign Ministry, the ship entered Freetown to replenish its stocks of water and to pick up British and West German scientists who were to join the research team. The 72-hour notice of entry required by the Sierra Leonean authorities was duly given, the foreign ministry said.

Why the Sierra Leonean impounded the ship is not clear. According to the Sierra Leonean High Commission in London, there was some suspicion that there were marines on board. This may explain the assertion, in the Soviet protest note, that "a strictly scientific vessel is being made out to be almost a military ship that had entered the territorial waters of Sierra Leone illegally". The attitude of the Freetown authorities, the note said, "gives reason for believing that this is a case of deliberate provocation", and the Soviets cautioned Sierra Leone that it should "seriously weigh up all the possible consequences to which its unfriendly steps with regard to the USSR could lead."

The *Akademik Mstislav Keldysh* was eventually released and has resumed research operations. Vera Rich

compounds (1–8 in the table), and to forbid export of these substances into 'war zones', although it will be some time before these new requirements will come into effect.

The dual-purpose nature of many chemical weapon-precursors makes international monitoring difficult. Controls were reportedly considered for isopropyl alcohol because it is a key ingredient in the mixing of binary chemical weapons. The new West German export regulations foresee mandatory licensing for the export of compounds 9 to 17 (see below), as is US practice for exports to Libya, Syria, Iran and Iraq.

The West German chemical industry organization Verband der Chemischen Industrie agreed last week to voluntary registration of exports of 23 additional compounds. The new regulations will for the first time require licences for export to 'sensitive areas' such as the Middle East. Steven Dickman

Protests as Bhopal settlement rebounds on government

New Delhi

THE long legal battle between the Indian government and the Union Carbide Corporation (UCC) of the United States came to an end last week when the Indian Supreme Court fixed an amount of \$470 million as compensation for victims of the Bhopal gas tragedy.

The court last week directed the US company to pay the Indian government \$425 million and its Indian subsidiary, Union Carbide India Limited (UCIL), the rupee equivalent of the balance of \$45 million as "final settlement of all claims, rights, and liabilities" arising out of the

the night of 2 December 1984, from a pesticide factory designed and built by UCC and operated by the UCIL. Soon after the tragedy, the Indian government took over the rights of its citizens and has been acting as the sole plaintiff in all suits against UCC — first in a New York district court, then in Bhopal and finally in the Supreme Court in New Delhi where the 38-month-long legal battle ended.

Although the verdict has been accepted by the Indian government, it has evoked extreme reactions in the press and in opposition political parties, which allege that the government has mishandled the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

An angry demonstrator pictured outside the Union Carbide plant in Bhopal following the "grossly inadequate" settlement just awarded by the Supreme Court.

world's worst industrial disaster. In fixing responsibility for the disaster, the court thereby indirectly exonerated Carbide from the charges of criminal negligence and violation of safety procedures at the pesticide plant.

The total compensation, to be paid before 23 March, is less than the \$600 million said to have been offered by UCC as an out-of-court settlement two years ago, and is about 15 per cent of what the Indian government had initially sought in damages.

The five judges of the constitution bench of the Supreme Court have acted in an unprecedented manner by announcing the compensation even before hearings were complete. The court was hearing arguments relating to interim compensation, so that its verdict on a "final settlement" was a surprise. The Indian government, which has blamed UCC's defective technology for the disaster, and which had sued the company for \$3,300 million in damages in a Bhopal trial court, apparently accepted the much lower figure to prevent the matter dragging on any longer.

According to official figures, 2,660 people died and between 30,000 and 40,000 were injured by the release of 40 tonnes of toxic methyl-isocyanate gas, on

case. One newspaper described the verdict as a "judicial let-down" and alleged that the level of compensation "is grossly inadequate, indeed laughable". Opposition parties, calling the court ruling "a betrayal of victims and a sell-out to the US multinational", have launched nationwide protests. The settlement issue is sure to be used as ammunition against Prime Minister Rajiv Gandhi's Congress party in the general elections next November.

Popular dissatisfaction over the verdict was also evident in Bhopal, where there was a protest march to the governor's office, and in New Delhi, where police had to be called to disperse violent civil rights groups outside the court.

The dust raised by the controversial court ruling will take time to settle, but the government faces the immediate problem of disbursing the money to surviving victims and relatives of those who died. There are 575,000 claimants and only half the claims have so far been scrutinized. India's track record of relief operations during floods or droughts has not been good, and the biggest question is whether the money due from UCC — the largest amount ever handled in a relief operation in India — will reach the victims.

K. S. Jayaraman

Generic drug-makers win telling point

London

GENERIC drug manufacturers in Britain have won the right to market out-of-patent drugs without themselves having to carry out toxicity tests and clinical trials. That is the effect of a decision last week of the law lords, the subcommittee of the House of Lords which functions as the nearest thing in Britain to a supreme court, which decided that information provided in confidence by companies seeking a product licence can also be used by the licensing authority when considering applications to produce generic copies of the drug.

The case turned on efforts by the United States-based pharmaceutical manufacturer Smith Kline and French (SKF) to resist generic copying of its successful anti-ulcer drug cimetidine (sold in Britain since 1976 as 'Tagamet'). Ironically, this is one of the drugs developed for SKF by last year's Nobel prize-winner Sir James Black.

Full patent protection on the drug expired last year; since then generic drug manufacturers have been seeking licences to produce copies of the drug. A directive from the Council of the European Communities says that a manufacturer can apply for a licence in such circumstances, and without having to carry out its own tests and trials, provided that it can demonstrate that its products are "essentially similar" to one already licensed.

SKF has argued that in order to demonstrate similarity, generic manufacturers must produce the results of tests and trials, and obtained a High Court ruling in its favour in February 1988. But the Appeal Court decided otherwise, whereupon SKF took its case to the House of Lords.

Last week, in their judgement, the law lords accused SKF of trying to "harass and obstruct" the licensing authority in order to prolong patent protection for its drug. But Cyril Beck, managing director of the generic drugs company Harris, says the legal arguments have extended SKF's monopoly on Tagamet by a year. He estimates that SKF sells £5 million worth of the drug every month.

Christine McGourty

Embryo research

Paris

Faced with divergent European laws and practices in the use of human embryos for research, the parliament of the Council of Europe has adopted a recommendation to set up an international body to monitor trends in this controversial field.

The council also looks forward to the possibility of formulating more uniform regulations among member nations, in an effort to prevent the emergence of 'genetic havens' in countries within Europe where laws are lax.

Peter Coles

US database link helps open up Japan to the West

Tokyo

JAPAN's Ministry of Education, Culture and Science (MESC) is poised to counter charges that the information flow between the West and Japan is mostly in one direction, by giving US researchers on-line access to its scientific databases. Computers at Tokyo's National Center for Science Information System (NCSIS) and the Washington headquarters of the US National Science Foundation (NSF) have just been linked through international telecommunication lines and will reach out to US researchers through NSF networks from April.

The new link, which will be extended to the United Kingdom and then Europe next year, complements efforts by the Science and Technology Agency to internationalize the massive databases it runs at the Japan Information Center of Science and Technology (JICST). But there is still some way to go to provide electronic satisfaction to US negotiators' demands for "comparable access" — including access to basic technological research in private companies — voiced during negotiation earlier this year of the US-Japan Science and Technology Agreement (see *Nature* 337, 396; 2 February 1989).

NCSIS's Professor Shoichiro Asano hopes that the first step, of providing access to MESC's academic databases, may make it possible for Japan to tackle the more difficult second and third steps of widening access to government research institutes and to research in private industry. Much of the latter research, he points out, is "invisible" even to Japanese eyes. The US link is being financed entirely by MESC, with technical back-up and administrative help from NSF. Asano is hopeful that this 'present' will demonstrate Japan's commitment to open communication.

For foreign researchers, the most valuable feature of the new link may be access to abstracts of academic conferences held in Japan. Abstracts are now often available in electronic form a few weeks after a conference takes place and long before a full scientific paper is published. Asano says that academic societies in the fields of electronics, computer and information sciences are making special efforts to ease foreign access by providing abstracts in English; researchers in the chemical and medical sciences are not far behind. Presentation and format will improve, he says, as evaluation comes in from foreign researchers.

The link also provides electronic mail, and access to the on-line academic library network and MESC's database of grants for scientific research. The ordinary user will, however, be restricted to the fraction of the database written in English, as special equipment (which has been installed at NSF headquarters) is needed to display Japanese characters. But information in English should help target material worth translating.

The language problem is not easily overcome. In late 1987, JICST began international on-line service by linking its own databases (which emphasize space and nuclear power) with the Chemical Abstracts Service in the United States and the Energy, Physics and Mathematics data-base (FIZ-Karlsruhe) in West Germany, but so far the flow of information has been mainly to Japan. The part of JICST data base of particular interest to foreigners concerns Japanese research; most of that is in Japanese and therefore inaccessible. JICST looks for a solution in machine translation, on which it is working hard. But a further problem may still remain: even when special projects, like that at George Mason University in Virginia, have given access to Japanese data, few US researchers have shown interest.

Alun Anderson

Europe's fast reactor plan inches forward

Bonn

FRANCE, West Germany, Britain and several other European countries have taken a further step in their plan to develop a fast reactor in collaboration with each other.

Representatives of the governments and their participating research institutes and industries last week signed the formal agreement meant to carry them forward to the next phase of the project — that of the detailed design of a 1,500 MW fast reactor, which should be completed by 1993. If built, the reactor would cost £2,000 million.

The new design will essentially be based on that of the French Superphénix 1, but will have either three or six coolant circuits and will be smaller in size. Superphénix 1 itself was restarted only last month, 18 months after it was closed on account of a leak from a liquid sodium circuit.

The three-year design programme will cost £190 million over three years, with France, West Germany and the United Kingdom each paying nearly a third. Belgium, Denmark, the Netherlands and Italy will make smaller contributions. John Collier, chairman of the UK Atomic Energy Authority, says that it would be possible for the reactor to be built and operating by 2000, provided that construc-

tion funds and an agreed site for the reactor could be found.

France, considered to be the leader in fast reactor development, is the most likely provider of a site, especially because of the political opposition expected in Britain and West Germany. By 2005, France will in any case need new sources of electricity generation to replace its earlier generations of nuclear reactors. But, in the past few years, the economic advantages of fast reactors have been challenged by the emergence of the spectral-shift pressurized-water reactor.

The continued support of the United Kingdom for the project is uncertain following the government's decision that its own fast-reactor development programme should be cut from £110 million a year to £10 million a year. The electricity supply industry, which is soon to be privatized, seems to be an unlikely alternative source of funds.

The future of West German participation is also uncertain. The demonstration fast reactor at Kalkar in North Rhine-Westphalia does not yet have an operating licence after 16 years of construction, 18 planning permits and DM7,000 million of investment.

Christine McGourty

Chinese satellites look to the future

London

CHINA's hopes of becoming a major launcher of space vehicles and satellites may have dimmed in recent years but have not been entirely extinguished. The difficulties of the US shuttle programme and the Ariane launcher had raised the prospect that China's 'Long-March-3' rocket might be used to launch satellites from overseas. But planners and experts at a meeting last month acknowledged that foreign competition and the restrictions imposed by the Coordinating Committee for East-West Trade Policy (COCOM) were a barrier to the commercial exploitation of Chinese launch facilities.

What the meeting decided is that, in the circumstances, China would have to offer a complete service, including sophisticated satellite technology, and not just launch facilities. Experts present took the view that the Chinese technology for retrieving satellites has a competitive advantage, especially if it is possible to develop a recoverable satellite for microgravity experiments, for which an important market is expected.

Meanwhile, China plans to concentrate on the use and development of communications and remote-sensing satellites. There are plans for a space station and a shuttle, but these are not expected to become realities this century.

Vera Rich

Truce on the US horizon for engineered organism releases

Washington

A LEADING group of US ecologists has stepped into the fracas over where to draw the line in regulating the release of genetically engineered organisms into the environment. The Ecological Society of America (ESA) last week proposed a scheme for determining the appropriate level of oversight for environmental releases, based on a "qualitative to semi-quantitative" estimate of the degree of ecological risk.

Establishing procedures for permitting field tests of genetically engineered organisms has been a tortuous process for the

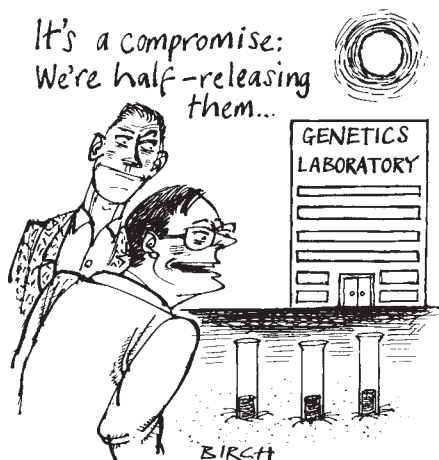
Biotechnology trade associations and environmental groups have been clamouring to register their objections to the proposals ever since. Last week, the EPA sought to rectify the situation by publicly announcing that the documents were available, and by defining a comment period until the middle of May.

The most contentious issues in the proposed rules have been how EPA should determine which organisms should be subject to review, how extensively the agency should control commercial research and development activities, and whether research carried out at universities on behalf of companies should also be monitored.

The Ecological Society's outline does not address these issues, but instead lays

out a sliding scale of scrutiny to be used on a case-by-case basis to assess the genetic changes made to the parent organism and the likely extent of changes in its behaviour in the environment.

The Ecological Society's plan was coordinated by James M. Tiedje of Michigan State University, who also served on the advisory panel for a report on the field-testing of genetically engineered organisms prepared last year by the US Office of Technology Assessment. Other environmental groups have expressed their approval of the plan, including such diverse organizations as the Conservation Foundation, the National Wildlife Federation and the Environmental Policy Institute/Friends of the Earth. Even Jeremy Rifkin of the Foundation on Economic Trends has added his approval: in its litigious tradition, the foundation has filed a petition through its attorney to force the EPA to consider the ESA's findings in rewriting its rules. Carol Ezzell



Environmental Protection Agency (EPA) and the Department of Agriculture — the two federal agencies given key regulatory jurisdiction within the Coordinated Framework for the Regulation of Biotechnology published in 1986. While the agencies have drawn criticism for not being sufficiently cautious in preventing the release of potentially disruptive or dangerous organisms, they have also been attacked for attempting to stifle the innovations of the biotechnology industry.

The EPA drew up a set of rules last summer setting out several controversial proposals under the Toxic Substances Control Act (TSCA) and the Federal Insecticide, Fungicide and Rodenticide Act, the two statutes EPA is adapting to cover biotechnology.

One would establish a network of local 'environmental biosafety committees' to parallel the National Institutes of Health's institutional biosafety committees, which review recombinant DNA experiments (see *Nature* 331, 107; 1988). Another would create a streamlined TSCA Experimental Release Application for approving limited commercial field tests.

The draft rules leaked out last autumn, during review by the Reagan administration, before they could be published in the *Federal Register* for public comment.

Research foundation loses out in fund-raising failures

Washington

WHILE the National Institutes of Health (NIH) are trying to encourage closer ties between their researchers and industry, at least one researcher has found the commercial road to be rocky.

In the spring of 1988, Robert Glazer of the National Cancer Research Institute received a \$500,000 grant from a pharmaceutical company for research on drug resistance. To make the money go further and avoid personal taxation, Glazer, with what he thought was the permission of his division director, established a foundation to administer the fund.

But in September last year, after the foundation had been established, NIH requested him to submit a formal application seeking approval for the foundation. One month later, the assistant director of the cancer institute recommended that the foundation be approved. But in November, NIH officials told Glazer that a more appropriate mechanism for using the money from the grant would be a Cooperative Research and Development Agreement (CRADA).

CRADAs were introduced last year under the Federal Technology Transfer Act of 1986 to promote collaboration between NIH researchers and industry (see *Nature* 336, 103; 1988). But Glazer's benefactor was uninterested in a CRADA, because the grant was not intended for commercial development purposes. Finally, in January, NIH rejected approval of Glazer's foundation.

Complicating the picture is a dubious scheme by Glazer to raise additional money for his foundation by means of a

sweepstake in which participation was offered publicly. Glazer contracted the services of an agency, Watson and Hughey, to run his fund-raising campaign. The firm has run campaigns for several cancer charity organizations in the United States.

Under the scheme, solicitations were mailed to several hundred thousand people informing them that they were winners in a \$5,000 sweepstake, and asking that they claim their winnings with a reply and a donation for the foundation. But the sweepstake has several drawbacks. Most 'winners' receive only 10 cents, the bulk of donations appears to be absorbed in mailing costs and very little goes to the client. Watson and Hughey are currently being investigated in several states for alleged breaches of consumer laws. But Glazer insists that their operation is legal.

NIH officials deny they were influenced in their decisions on Glazer's foundation by the sweepstake scheme, which came to light after the National Cancer Institute's own in-house journal published a report on Watson and Hughey in January (*Journal of the National Cancer Institute* 81, 18 January, 1989).

Glazer has agreed to stop the fund-raising and transfer his grant to another foundation associated with NIH. But he is furious that the cancer institute's journal revealed his sweepstake scheme after he had agreed to abandon it. "I haven't made a dime" out of the sweepstakes, Glazer says, who says he is considering taking legal action for violation of his privacy. David Swinbanks

US – Soviet collaboration on biosatellite is a success

Mountain View, California

US and Soviet scientists last week gathered at the NASA (National Aeronautics and Space Administration) Ames Research Center to announce the results of their most successful collaborative effort so far in the life sciences, and to make plans for future collaboration. The most recent mission, *Cosmos 1887*, has paid off, with new insights into the physiological changes induced by prolonged weightlessness.

NASA has been looking for opportunities to raise its profile in the life sciences. Several agency surveys in the last year have stressed the need for expanded bio-

medical research, although funds have not been forthcoming from Congress. The research, however, is "absolutely essential . . . for long duration human exploration, either back to the Moon, or on to Mars", according to Noel Hinners, NASA's chief scientist.

The United States has participated in six Soviet biosatellite missions since 1975, under the terms of a 1971 bilateral agreement for scientific collaboration in space. The programme has been a bargain for NASA, as all flight costs are paid by the Soviet Union. NASA spent \$1 million on ground-based laboratory analysis for *Cosmos 1887*, but has now invited Soviet

participation in the first space shuttle mission dedicated to the life sciences, due to fly in June 1990.

Cosmos 1887 orbited the Earth for two weeks in October 1987, carrying two rhesus monkeys, 10 male rats and a collection of fish, amphibians, birds and mammalian cells in tissue culture. It achieved notoriety when one of the monkeys on board freed his left arm and tampered with experiments within his reach. The Soviet director of the *Cosmos* biosatellite programme, Eugene Ilyin, shed some light on that incident, attributing the monkey's misbehaviour to a failure in the equipment that was supplying its food.

A failure in the craft's braking system caused it to land 3,000 miles off course, in eastern Siberia, and resulted in a 2-day delay in tissue analysis. Nevertheless, the mission provided the most detailed data yet on the physiological effects of weightlessness. Among the findings were a 40 per cent shrinkage in rat muscle fibres, accompanied by a 60 per cent drop in contractile protein. The rats' bones showed a 30 per cent decrease in their ability to resist bending and breaking, apparently due to a redistribution of minerals away from the midshaft of the bone, said Richard Grindeland, biospecimen programme manager for NASA.

The rats also showed a 50 per cent reduction in their circulating levels of growth hormone, which normally plays a role in the maintenance of both bone and muscle. Grindeland says this may mean that the bone and muscle wastage common to space travellers and thought to be caused by lack of use may be caused partially by hormonal changes.

Neurological studies of the rhesus monkeys have found an increased excitability of the vestibular nerve, which could explain the 'space sickness' or nausea that astronauts and cosmonauts often experience, said Ilyin. Weightlessness was also found to take a toll on the immune system. Adrian Mandel of NASA-Ames said the rats showed a loss of helper T cells, leading to a reversed T4 to T8 cell ratio, reminiscent of that found in AIDS. Because the animals were killed within days of the spacecraft's return, the course of recovery from the immune suppression is not known.

Grindeland said none of the results obtained so far would rule out the possibility of long-term space travel, although they point to the need for further research, especially into whether the effects observed with rats occur in monkeys as well. Muscle biopsies will be taken from the monkeys on the next *Cosmos* mission. Grindeland said he expects that further research will guide the design of special exercise and pharmacological regimens to counteract the deleterious effects of weightlessness.

Marcia Barinaga

Europeans join forces to design a bigger and better chip

Munich

WEST German Research Minister Heinz Riesenhuber last week announced plans for six European countries to cooperate with three large electronics companies in developing a new generation of 64-megabit silicon memory chips.

The project represents European co-operation in microelectronics on an unprecedented scale — the estimated budget is DM8,000 million over seven years. It follows a bilateral West German–Dutch collaboration to design and produce a 4-megabit chip. The West German contribution would be about DM3,000 million, one-third of which would come out of the budget of the Research and Technology Ministry (BMFT), the rest from industry.

Riesenhuber warned of the "danger" that would threaten European industry if it were to become dependent on its foreign competitors for memory chips. He urged the European Community to act quickly in approving the project, named JESSI for Joint European Submicron Silicon.

JESSI is divided into linked sections. The largest share of the proposed budget — nearly one-third — would be spent on the development of 'designer chips' for specific applications and of tools for computer-aided design. Basic research will focus on developing design and integration methods for the new chips. Under the plan, logic circuits and production engineering will be developed specifically for the new chips, as will be the large-scale production facilities such as lithographic equipment and 'clean rooms'.

The six nations involved are West Germany, Britain, France, Belgium, the Netherlands and Italy. West Germany is the first country to give its official approval

to the project. The companies involved are the West German microelectronics giant Siemens, its Dutch competitor Philips, and the French–Italian conglomerate SGS-Thomson. Siemens and Philips have been working together to produce the 4-megabit chip.

The European Community's response to the project is critical, say sources in West German government and industry. Organizers hope that the Community will agree that JESSI be run as a EUREKA project, which would allow the individual member states and companies to make independent decisions. There would be disastrous consequences, say the participants, if the Community were to set up a competing research programme under its own auspices.

An official in Brussels who insisted on anonymity offered reassurance to the worried West Germans: "there is no misunderstanding" of the importance of JESSI, and the Community plans to work with the six member states "like seven fingers on one hand". But the official could not say when the decision would be made.

Riesenhuber's early announcement is one way of staying one jump ahead of West German critics of JESSI. Criticism is likely to come from those who oppose spending public money on applied research or on research that will benefit large companies. There is a strong lobby in West Germany for companies of small and medium size, which form the backbone of West German industry.

Riesenhuber emphasized that small industry will benefit at least indirectly from JESSI, but admitted that details about how small companies can participate have yet to be worked out. **Steven Dickman**

Byelorussia still alarmed by the effects of Chernobyl fallout

London

GLASNOST promises to bring some relief to the many in Byelorussia who are alarmed about their continued exposure to fallout from the Chernobyl accident. After months during which public demands for more information have been dismissed by officials as "radiophobia", Byelorussians have now been promised a chance to put their questions and anxieties to panels of

Mogilev and Gomel provinces are to be evacuated because of the persistence of caesium-137 there.

The day after this announcement, a first public meeting on Chernobyl was held at the House of the Trade Unions at Minsk, the capital of Byelorussia. The highlight of the meeting was the production of detailed maps of the contaminated areas which, together with a transcript of the

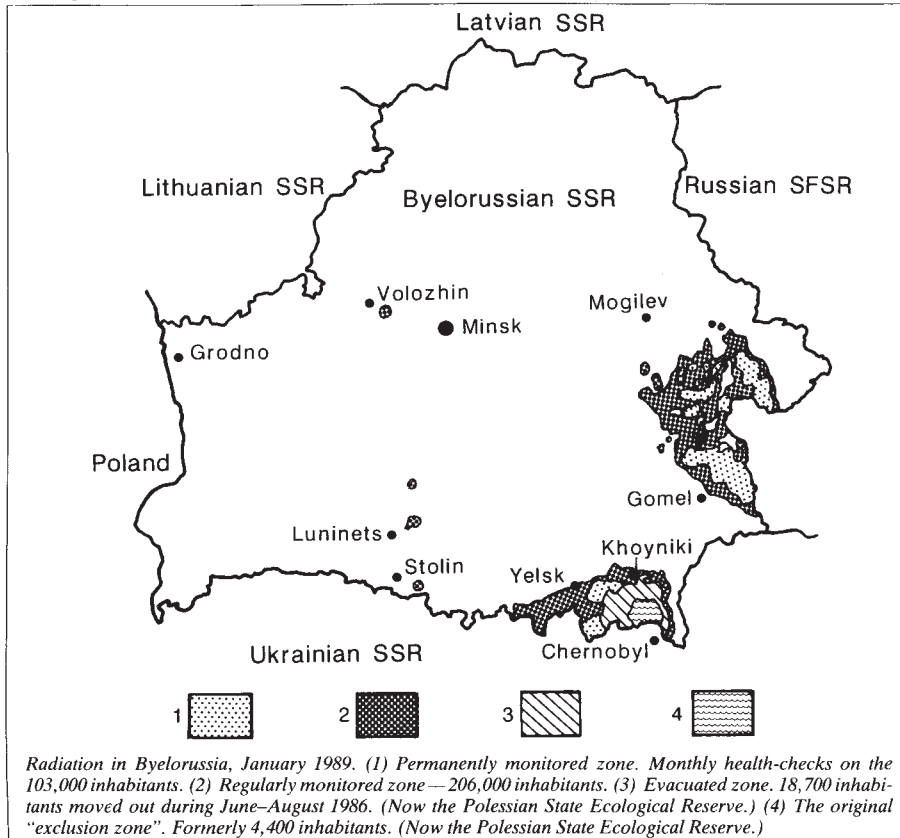
rem in the third.

Bur'iak also said that the dose received by the population in the monitored zone of Byelorussia over the whole of the period since the accident amounted to 9.0 rem, compared with 3.3 rem for the rest of Byelorussia. But he also acknowledged that 48 pensioners who had ignored official advice and returned to the homes from which they had been evacuated had received doses exceeding the limits, while insisting that "none of those who were evacuated received the limit dose of 75 rem".

Although the opening statements by officials at the meeting suggested that the radiation situation in Byelorussia is well in hand, questions from the floor occasionally produced evidence of muddles and mistakes. Thus it was acknowledged that the civil defence manuals had not foreseen the need to evacuate rural areas, while there have been delays in providing the promised tractors with sealed cabs for working contaminated land more safely. Some participants at the meeting, dissatisfied with assurances that food reaching the shops is safe, demanded radiation monitors so as to be able to check for themselves.

Many speakers at the meeting were sharply critical of the official approach to the clean-up operations, suggesting for example that it would be more economical to let contaminated land lie fallow than to attempt to produce "safe" grain and meat from it. Complaints that illnesses have been more common in the region since the accident were met, at the meeting in Minsk, with the reply that there was an influenza epidemic and that medical monitoring brings more minor illnesses to light. But why, one questioner asked, have so many medical personnel, who must know the real risks, quit the region? Plainly, many more public meetings will be needed before the population is content.

Vera Rich



experts and scientists.

Although Chernobyl is in the Ukrainian republic, the prevailing winds meant that Byelorussia, to the north, caught much of the fallout from the Chernobyl accident nearly three years ago. For several months, there have been indications that the radiation situation in the republic is more serious than had at first been allowed.

Last October, for example, the Prime Minister of Byelorussia, Michal Kovaleu, told the Supreme Soviet of the Soviet Union that "the effect of the Chernobyl disaster in the republic does not decrease". More recently, he told a *Pravda* reporter that a fifth of the agricultural land in Byelorussia is contaminated.

On 24 January this year, the Supreme Soviet voted an extra 243 million rubles for expenditure in the republic related to Chernobyl, while on 1 February it was announced that a further 20 villages in the

proceedings, have since been reproduced in the two leading dailies of the republic.

Among other things, the maps show that 103,000 people live in areas where monthly health checks are required, and that there are 206,000 people in a zone where "regular" checks have been ordered. In the more seriously contaminated zones, no vegetables or fruit may be grown, but local people receive a 25 per cent salary bonus in compensation for the cost of buying food from elsewhere.

There is some confusion about the radiation exposure allowed. A lifetime of exposure of 35 rem has been adopted as a general limit for the population, and evacuation plans are determined by estimates of whether this limit will be exceeded. At the public meeting in Minsk, a deputy minister of health, V. M. Bur'iak, said the authorities are working with limits of 10 rem in the first year of exposure, 3.0 rem in the second and 2.5

More West German university investment?

Bonn

THE federal government is prepared to spend more on universities if it is joined by the *Länder*, the West German Education Minister Jürgen Möllemann announced on 10 February. Möllemann suggested an increase of DM600 million a year in the federal budget for university construction, with half of that amount coming from the *Länder*. The budget now stands at DM2,000 million.

The DM2,000 million "emergency programme" announced in December to relieve chronically overcrowded university courses (see *Nature* 336, 704; 1988) will take effect as soon as it is approved by the heads of the *Länder*. Steven Dickman

Reprints: more for than against

SIR—Current awareness publications such as the *Current Contents* series play an important role in disseminating information about where to find published research. But there can be no doubt that these information tools contribute to the problem described by Ivor Smith (*Nature* **336**, 708; 1988), the indiscriminate use of reprint requests.

In 1988, I published two articles dealing with the history of molecular biology; one article described the discovery of RNA splicing and the other gave and discussed a somewhat tongue-in-cheek list of significant discoveries in molecular biology over the past fifty years. The articles appeared in sister review journals — *Trends in Biochemical Sciences* (**13**, 110–113; 1988) and *Trends in Biotechnology* (**6**, 234–243; 1988) respectively — but whereas the former article generated more than 250 reprint requests, the latter attracted only one-fiftieth of that number. It might be argued that the splicing article was simply a better article, but I ascribe the difference in my mailbag to another factor. *Trends in Biochemical Sciences* is listed in *Current Contents Life Sciences*, while *Trends in Biotechnology* is not.

JAN A. WITKOWSKI

Banbury Center,
Cold Spring Harbor Laboratory,
PO Box 534,
Cold Spring Harbor,
New York 11724, USA

SIR—Ivor Smith's criticism of the reprint, on the basis of costs and inconvenience alone, ignores the real merits of this form of scholarly communication. Reprints represent a direct, shorthand way for scholars to get in touch with others doing similar work. This personal contact is important, as it may lead to further communications, or perhaps later collaboration on some research. There is also the psychological value of knowing that others are interested in one's work. This immediate, favourable feedback undoubtedly inspires many authors to pursue their work with renewed vigour.

Contrary to Smith's belief, thousands of researchers around the world, especially in developing countries, do not have easy access to the published literature, so photocopying is a moot issue. For others, reprints are the only convenient medium for developing good subject collections, in cases where the literature is widely dispersed in various fields or simply unavailable because of inadequate distribution of the original journals.

Finally, reprints, printed on quality paper, are an excellent archival medium for personal libraries and ultimately research libraries which receive these materials as part of gift collections.

Arguably, some abuse the privilege of requesting reprints, but the case for reprints is too great to stop their use altogether.

JOHN H. SANDY

Natural Resources Research
Institute,
University of Minnesota, Duluth,
Duluth, Minnesota 55811, USA

SIR—I take exception to Ivor Smith's suggestion that photocopying replace the reprint. There are at least six reasons why reprints should survive.

(1) Many young faculty send reprints to more established colleagues as a simple but powerful way of making professional contacts.

(2) Many academic search committees require that applicants provide copies or reprints of recent publications. Young people seeking a permanent position may apply for dozens of jobs a year. They cannot be expected to bear the burden of photocopying 100 or more pages each time.

(3) Many of us do not have grants to cover the considerable cost of photocopying the large number of papers we need to read each week.

(4) Many university departments have photocopiers, but the smaller institutions do not permit unlimited photocopying.

(5) Most small universities do not have library subscriptions to all the appropriate journals. Consequently, we have to keep abreast of the literature by scanning *Current Contents*, *Current Advances*, *Biological Abstracts* and so on. Thus, the only way to obtain many titles is to request a reprint from the author.

(6) Many journals provide at least 50 free reprints. Why not save one's photocopying money, use it to buy extra reprints, and be kind enough to honour reprint requests?

At the very least, one could provide reprints to people one has never heard of (and who are therefore likely to be students or otherwise needy), to people at poorer institutions or departments and to people one happens to know are struggling financially.

CHRISTOPHER P. DUNN

Biology Department,
Ball State University,
Muncie, Indiana 47306, USA

No closure

SIR—A. J. Southward (*Nature* **337**, 202; 1989) claims that the Natural Environment Research Council (NERC) "is proposing to run down" the Biological Records Centre at Monks Wood, on the basis of an article in *The Biologist* (whose author is identified by the pseudonym

"Peccavi")!

It is not NERC's intention to run down the Biological Records Centre (BRC). NERC fully recognizes the importance of biological recording and is actually strengthening the capability at Monks Wood by forming a new Environmental Information Centre. This will bring together the BRC and staff involved in remote sensing and data handling and will involve the transfer of additional staff to Monks Wood.

Furthermore, NERC is convening a meeting to discuss the implementation of the recommendations of the recent Linnean Society report *Biological Survey: Need and Network*, which should lead to improvements in this field.

P. B. TINKER

Natural Environment Research Council,
Terrestrial and Freshwater Services,
Polaris House, North Star Avenue,
Swindon SN2 1EU, UK

Science funding

SIR—You are right to warn against complacency about science funding ("Safeguarding British research", *Nature* **337**, 291; 1989). However, there is a note of complacency in your analysis of the problem. You suggest that if only the government would understand the economic usefulness of basic research, it would provide adequate funding. In addition, if only young people would appreciate the excitement of a research profession, academic science would be restored. Both premises are questionable. Rather than addressing only the avowed economic imperatives for government science policy, your article should also have examined the hidden political reasons behind the underfunding of academic science.

Academics have traditionally had the freedom not only to pursue novel scientific theories but also to consider and act on the wider implications of their research and of world events. As illustrated by your article, underfunding science ties up scientists' political energies in defending their economic usefulness. Furthermore, the current scramble for grants, meagre salaries and scarce university positions in the United Kingdom (which I recently left), selects against young people with interests wider than their research profession. The government, by simultaneously removing the financial security and career prospects of university positions, and undermining the social image and status of academics, is silencing a traditionally liberal and outspoken section of society.

The serious and lasting damage is indeed that done in people's minds. It is that we no longer dare to ask for intellectual freedom.

MARIE ROSE VAN SCHRAVENDIJK
4277 Pomona Avenue,
Palo Alto, California 94306–4313, USA

Einstein's much delayed little joke

When Einstein invented the cosmological constant in 1917, and promptly disowned it, he cannot have known that the concept would be bothering his successors more than 70 years later.

EINSTEIN had very few reasons to be ashamed of any of his work, but his invention of what is called the cosmological constant seems to have been one of them. And it is a curious business. Having developed the theory of gravitation called the General Theory of Relativity, Einstein embarked in 1917 on an attempt to apply his equations to the Universe as a whole. On the principle that the motions of the most distant objects in the Universe — the first galaxies to be recognized as such were being measured at the time — are uniformly much less than the velocity of light, Einstein set out to find solutions of his equations that might be taken to represent a static Universe, but could not.

So, with a daring that must have seemed natural to one who had just used pure thought to put the then laws of physics on a rational foundation, he invented the cosmological constant and promptly reached a solution of the desired form — a spherical fixed universe filled with matter of uniform density. But within a few months, Einstein had withdrawn the notion of the cosmological constant.

Legend has it that Einstein's disillusion sprang from Hubble's discovery that the Universe is expanding (so that the search for a static solution would have seemed a wild goose chase), but Steven Weinberg, writing from the University of Texas at Austin, tells a different story in an elegant review article just published (*Rev. Mod. Phys.* **61**, 1; 1989).

Serious accounts of the dilemma make it plain that the difficulty that gave Einstein pause was de Sitter's demonstration a few months later that there is a static solution of Einstein's equations with the uncomfortable property that the Universe it represents is entirely devoid of matter. This may have seemed a body blow, for it cast doubt on Einstein's guiding principle that mechanical inertia is a consequence of the distribution of mass within the Universe.

In the event, as Weinberg puts it, the need for the cosmological constant was entirely undermined by the emergence in the 1920s of the family of cosmologies due to A. Friedman, which allowed for an expanding or a contracting Universe simply by the appropriate choice of a parameter. Yet the irony remains that, despite Einstein's disavowal of his own invention, the question of whether the cosmological constant exists, and is different from zero, and of what precisely it

may mean, is more vividly alive now than sixty years ago.

The bare bones of the issue are easily understood. Einstein's picture of the geometry of space-time turns on the set of 16 coefficients defining the relativistic equivalent of distance (or time) by means of a quadratic form called a metric; only 10 of these coefficients are independent, because the form is symmetric, but general relativity differs from special relativity in that the coefficients are functions of position. The original field equations relate these coefficients at any place-time to the curvature at the same place-time, which is a function of the same coefficients. (This is why the equations are not linear and are thus not soluble systematically.)

The physics of general relativity is embodied in the way in which the metric coefficients are related to the density of energy (which includes mass) and momentum in the system. The effect of Einstein's introduction of his cosmological constant is dimensionally and in every other way equivalent to the arbitrary addition of energy and momentum to empty space, in quantities determined by the size of the constant. That explains why the constant should have turned an expanding into a static universe. In the distant early 1920s, it is understandable that Einstein should first have taken the liberty of introducing the cosmological constant and then, when it seemed no longer necessary, the liberty of discarding it.

Quantum mechanics has put a stop to that. That is Weinberg's starting point. The essence of his case is that, whatever Einstein's pros and cons may have been, nobody can now hide from the truth that even empty space is filled with energy — to say the least of it, the zero-point energy of all the oscillators corresponding to radiation of different frequency with which even a vacuum must be filled but, more tangibly perhaps, because pairs of particles (such as electrons and positrons) can be created out of nothing anywhere and any time by what are called the fluctuations of the vacuum.

From this observation springs the dilemma that now perplexes people. If the effect of the unavoidable fluctuations of the vacuum are logically equivalent to a cosmological constant different from zero, and if the observation of the real Universe shows that Einstein's cosmological constant is unnecessary, then the

underlying equations of general relativity must include a 'cosmological constant' whose value is exactly cancelled by that simulated by the fluctuations of the vacuum. And the cancellation must be extraordinarily exact.

The net effect of the 'real' and 'simulated' constants can be related to Hubble's constant, which determines the rate at which the Universe is expanding, and which serves as an upper bound, while the value of the constant simulated by vacuum fluctuations can be estimated from the properties of quantum systems. Weinberg concludes that the two lines of argument are consistent only if the true and simulated values of the cosmological constant cancel each other out to 118 decimal places (although, on another estimate of the quantum fluctuations, he estimates that the cancellation may extend to only 41 decimal places). For practical purposes, this is simply another way of saying that the numbers are equal.

Weinberg's objective is to consider how the extraordinary coincidence implied by these estimates can be rationally explained. He flirts with several possibilities, only to dismiss them. The supersymmetric world in which fermions (electrons, say) and bosons (say photons) are identical might arrange for the cancellation if the supersymmetric world were like the real world (which it is not). The superstring theory of particle fields (in which six of ten dimensions are made to disappear, giving entities which are elastic strings in ten dimensions, the properties of the particles we know in only four) appears to have offered hope at first, but to have proved a disappointment.

The anthropic view, that people are able to observe the stars and galaxies only if the Universe in which they find themselves is one in which the cancellation is exact is plainly seductive, but Weinberg eventually dismisses it in each of half a dozen varieties. His own clear favourite is that the cancellation will spring naturally from the attempts being made to develop a quantum theory of gravity, but there may be as many opinions as there are people qualified to write about the problem.

Meanwhile, it is clear that Einstein was even more perceptive than he knew. It cannot be often that the notions a person discards turn out to be crucial to a conundrum that perplexes his successors the best part of a century later. **John Maddox**

Oxide superconductors

A touchstone for theories

T.M. Rice

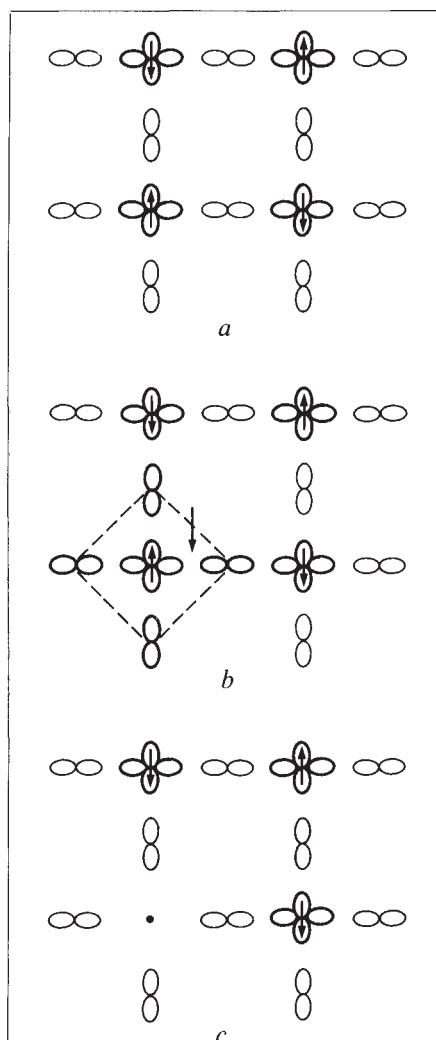
THE observation by Tranquada *et al.*, reported on page 720 of this issue¹, of singly charged copper ions in the superconductor $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ confirms the suggestion made in *Nature* last month by Tokura and co-workers² that this material is an electron-doped high-temperature cuprate superconductor. Tokura *et al.*² based their assignment on a straightforward chemical analysis, backed up by a measurement of the sign of the carriers through the Hall effect. The new experiments assure us that there was no hidden subtlety in the chemistry or structure and require us to face up to the theoretical implications of electron doping. In a News and Views article³ accompanying the report of Tokura *et al.*², Emery discussed how the new electron-doped superconductors could be used to distinguish theories of high-temperature superconductivity. However, in my view these experiments are a more decisive test of which theory correctly describes the mechanism than any of these authors has suggested.

To understand why, it is helpful to start with the electronic structure of the CuO_2 planes which are the key element of all cuprate superconductors. It is customary to measure formal valence relative to the closed-shell configuration Cu^+ (with 10 electrons in the Cu 3d valence orbitals; $3d^{10}$) on copper sites and O^{2-} (or $2p^6$) on oxygen sites. The undoped parent materials La_2CuO_4 (or Nd_2CuO_4) have formal valences of Cu^{2+} and O^{2-} , in which one hole has been added to (one electron removed from) the copper sites. These holes do not move freely but are localized on individual copper sites by a strong on-site Coulomb interaction (see *a* in the figure) and these materials are insulators. The $S=1/2$ spins of the holes are free but they order anti-ferromagnetically (alternate spins anti-parallel) owing to the 'super-exchange' interaction between them.

Superconductors are formed when these materials are doped or, in other words, extra electrons are added or removed from the CuO_2 planes by substituting ions of different valence on the out-of-plane lanthanum or neodymium sites. Until now all the cuprate superconductors were hole doped — more holes were added to the CuO_2 planes by replacing some La^{3+} by Ba^{2+} . It is energetically favourable for these extra holes to go primarily on the O-sites as emphasized first by Emery⁴. The resulting configuration (*b* in the figure) has two holes on a CuO_4 square. Electron doping was achieved by Tokura *et al.*² by substituting

Ce^{4+} for Nd^{3+} . This removes holes and leads to a local electronic structure which is a closed-shell configuration (*c* in the figure). This causes a copper site to revert to the Cu^+ configuration directly observed by Tranquada *et al.*¹. Why do these two seemingly different configurations in the end lead to the same highly unusual superconducting state?

As mentioned by Emery³, the theories



The hole orbits in the CuO_2 planes. Electrons removed from the closed shell configurations Cu^+ (or $3d^{10}$) on Cu sites and O^{2-} (or $2p^6$) on O sites give hole orbits. *a*, The undoped CuO_2 plane. The configuration is Cu^{2+} with one hole in a $3d_{x^2-y^2}$ orbit which admixes onto the O sites by covalency. *b*, Hole doping by adding an extra hole. In the t - J model, this extra hole is bound into a local singlet configuration on a CuO_4 square⁷. Electron transfer allows the singlet to move from square to square. *c*, Electron doping removes a hole creating a Cu^+ configuration as observed by Tranquada *et al.*¹.

of high-temperature superconductivity can be divided into two schools of thought. One goes back to the first theoretical paper written by Anderson⁵ very soon after the initial discovery of the effect by Bednorz and Müller. He proposed then that the cuprate superconductors were a new type of electronic system. The simplest prototypical model which could describe this system was made up of the $S=1/2$ spins on the Cu^{2+} sites coupled by a nearest-neighbour Heisenberg exchange (J) term and singlets ($S=0$) representing the squares with 2 holes. The strong hybridization (mixing of character) between copper and oxygen orbitals is the source of the singlet binding^{6,7}. These singlets carry an extra positive charge and are able to hop from square to square by the exchange of an electron (described in the theory by a matrix element t).

This model is now generally known as the t - J model. The key idea in this school of thought is not that the t - J model is a completely accurate representation of the cuprates, but that it contains the essential ingredients that define these new electronic systems. For example, the behaviour of magnetic insulators is usually discussed in textbooks in terms of Ising and Heisenberg spin models. Such models are simplifications of the real spin interactions in almost all materials, but the omitted terms do not influence the qualitative behaviour of the models.

All models of this class have a transition from a paramagnetic (statistically orientated spins) to a magnetically ordered phase as they are cooled through a critical temperature, which depends linearly on the strength of the exchange interactions. The study of the simplest models gives us a fairly complete understanding of the magnetic transition. A necessary consequence of this point of view is that any material which can be reduced to the same essential form must also have the same type of transition. Tranquada *et al.*¹ in this issue prove the existence of Cu^+ charge carriers which have a filled d shell and are indisputably singlets, so it is a clear requirement for this school that they also should have a transition to superconductivity, as indeed they do.

The second school has focused on specific details of the cuprates. For the hole-doped cuprates, Emery^{4,8} and others argued against binding of the holes in local singlet configurations, so that there are new degrees of freedom introduced. In this school the superconductivity transition is described as a variant on the Bardeen-Cooper-Schrieffer model which describes conventional superconductors. The cuprates are distinguished by having special electronic polarizabilities (from excitations on the oxygen sites with holes, for example), magnetically induced interactions or phonon-induced interactions which either separately or together

cause a strong attraction between the extra holes.

Because there are many metallic compounds that never superconduct, there have to be very exceptional circumstances to distinguish the cuprates from these. From this point of view it is far from clear that these circumstances can be the same when one changes the relevant orbitals by passing from hole to electron doping: the relevant interactions should all change. For example, a Cu^+ ion surrounded by O^{2-} ions is locally a filled-shell configuration and therefore only weakly polarizable, unlike the two-hole configuration (b in the figure).

Another interesting consequence of the experiments of Tranquada *et al.*¹ is that they favour a limit in which there are strong spatial correlations between the electrons, as shown by the separate Cu^+ and Cu^{2+} spectral lines observed. Only if the transfer rate (defined as $\hbar/t$, where $\hbar$ is Planck's constant) between Cu^+ and Cu^{2+} is slower than the electronic relaxation times, are these configurations clearly separated. Therefore, although one may be able to approach the superconducting state from intermediate or weak coupling

as Schrieffer *et al.*⁹ and Friedel¹⁰ propose, the electron-doped materials, at least, are closer to the strong coupling limit.

Despite two years of concerted study there is still no reliable theory of the superconducting transition temperature in the t - J model nor indeed any definitive proof that it does or does not exist in this model. The electron-doped cuprates offer us a more clear-cut example of materials that should be in the class of t - J models and therefore will encourage the continued attempts to solve this difficult many-body problem. □

1. Tranquada, J.M., Heald, S.M., Moodenbaugh, A.R., Liang, G. & Croft, M. *Nature* **337**, 720–721 (1989).
2. Tokura, Y., Takagi, H. & Uchida, S. *Nature* **337**, 345–347 (1989).
3. Emery, V.J. *Nature* **337**, 306 (1989).
4. Emery, V.J. *Phys. Rev. Lett.* **58**, 2794–2797 (1987).
5. Anderson, P.W. *Science* **235**, 1196–1198 (1987).
6. Anderson, P.W. in *Frontiers and Borderlines in Many-Particle Physics* (eds Broglia, R. A. & Schrieffer, J.R.) 1–37 (North-Holland, Amsterdam, 1988).
7. Zhang, F.C. & Rice, T.M. *Phys. Rev. B* **37**, 3759–3761 (1988).
8. Emery, V.J. & Reiter, G. *Phys. Rev. B* **38**, 11938–11941 (1988).
9. Schrieffer, J.R., Wen, X.G. & Zhang, S.C. *Physica C* **153–155**, 21–25 (1988).
10. Friedel, J. *Physica C* **153–155**, 1610–1616 (1988).

T.M. Rice is at the Institut für Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.

Brain diseases

Ubiquitous variations in nerves

Jean-Marc Gallo and Brian H. Anderton

ALL neurodegenerative diseases, the most common of which is Alzheimer's disease (see next page), are characterized by degeneration and eventual death of nerve cells, and in some of these diseases the vulnerable neurons contain characteristic fibrous structures or inclusion bodies. The fibrous inclusions contain elements derived from the cytoskeleton, and the current knowledge of their composition in various neurodegenerative diseases is summarized in the table. It is hoped that identification of the complete chemical composition of the inclusion bodies will elucidate the biochemistry of nerve cell death in these diseases. This hope has now come nearer to realization with the discovery that there may be a common sequence of events affecting cytoplasmic organization that accompanies and possibly results in the demise of neurons.

Despite the morphological and antigenic differences of the inclusion bodies in various neurodegenerative diseases, it has become increasingly clear that most of them share one component: ubiquitin, a 76-amino-acid polypeptide involved in the covalent modification of proteins. Ubiquitin is highly conserved — there is a difference of only three amino acids between the ubiquitin of yeast and of man — and it is found in all eukaryotic cells so far examined. It is synthesized either as

polyubiquitin or as fusion proteins with other non-ubiquitin sequences; both pre-ubiquitin forms are rapidly processed into free ubiquitin. Nevertheless, about half of the ubiquitin in cells exists as conjugates with target proteins through isopeptide bonds formed between the carboxy-terminal glycine residue of ubiquitin and the ϵ -amino side-chain of lysine residues of the target proteins.

Mori *et al.*¹ identified ubiquitin as a component of neurofibrillary tangles in

Alzheimer's disease by using specific antibodies together with amino-acid sequencing of peptides of the paired helical filaments that comprise the tangles. Since then, other groups have shown that most of the inclusions in Alzheimer's and other neurodegenerative diseases stain with antibodies against ubiquitin (see table)^{2–6}. These ubiquitin accumulations are so characteristic of degenerating neurons that ubiquitin-positive dense deposits and skein-like structures have now been discovered in the susceptible anterior horn cells of the spinal cord in motor neuron disease^{7,8}. Ubiquitin has also been observed in Mallory bodies, cellular inclusions not related to neurodegeneration but occurring in alcoholic liver disease⁹ and apparently derived from the hepatocyte cytoskeleton.

Although some monoclonal antibodies against ubiquitin stain the microtubule network of cultured cells⁹, ubiquitin is not usually found conjugated to the cytoskeletal proteins from which the inclusion bodies are derived. Ubiquitin accumulations are therefore probably related to the process of neurodegeneration rather than coincidentally trapped with cytoskeletal components during the formation of inclusion bodies.

The best-known function of ubiquitin is that it marks proteins committed to rapid degradation by an ATP-dependent pathway (see ref. 10 for a review). A protein needs to carry specific amino-acid sequences or amino-terminal signals to be recognized by the ubiquitin-mediated proteolysis system. Some altered or damaged proteins may display such signals and are therefore recognized as 'abnormal', inducing them to become ubiquitinated and subsequently proteolysed.

The role of ubiquitin as a cellular scavenger is an attractive explanation for its accumulation in neuronal inclusions. The presence of ubiquitin could reflect an overloading of the proteolysis system by abnormal or damaged proteins, including

Ubiquitination of neuronal inclusions in the most common neurological disorders

Disorder	Inclusion	Immunoreactivity	Ubiquitin	Ref.
Alzheimer's disease	Neurofibrillary tangles of paired helical filaments	Tau, fragments of neurofilament heavy and middle chains	+++	1–4
	Hirano bodies	Actin, tropomyosin, tau	—	5
	Granulovacuolar degeneration	Tubulin, neurofilament heavy chain	Some +	3
Parkinson's disease	Lewy bodies	Neurofilaments, tubulin	+++	5,6
Pick's disease	Pick bodies	Tau, neurofilament heavy and middle chains	+++	5,6
Progressive supranuclear palsy	Neurofibrillary tangles of straight filaments	Tau	Some +	4,5
Motor neuron disease	Dense and skein-like inclusions	?	+++	7,8

Tau, heavy and middle chains of neurofilaments, actin and tropomyosin are all protein constituents of the cytoskeleton. Plus signs indicate the degree of ubiquitination.

cytoskeletal proteins. Indeed, the cytoskeleton is profoundly modified in Alzheimer's disease: microtubules and neurofilaments are greatly reduced in neurons containing paired-helical filaments and tau is expressed in an abnormal phosphorylation state. Alternatively, the degradation system itself could be damaged so that target proteins are improperly ubiquitinated and not recognized by specific proteases, or the proteases themselves could be inactivated. But these hypotheses are still speculative because a ubiquitinated protein is not necessarily 'abnormal'. Indeed, ubiquitin has been found bound to proteins not committed to enhanced degradation: ubiquitinated proteins include histones H2A and H2B, actin in *Drosophila* flight muscle and some membrane receptors. The diversity of the proteins that can be conjugated to ubiquitin indicates that ubiquitination has many cellular functions, but there are few indications of its molecular form in cells and its physiological role in neurodegeneration.

The significance of ubiquitin in degenerating neurons will probably be understood when the ubiquitin moieties in paired-helical filaments and other inclusion bodies are characterized. Once putative protein targets for ubiquitination have been identified, it will be important to determine how these proteins become conjugated to ubiquitin, possibly by studying cellular models expressing the target proteins.

Indeed, ubiquitin is one of the group of proteins called heat-shock proteins which are induced in cells in response to stresses such as thermal shock, heavy metals, oxidants and amino-acid analogues. In conditions of stress, chicken fibroblasts increase their level of ubiquitin conjugates, and during recovery from stress they enhance their overall protein degradation¹¹. The study *in vivo* of the stress response to chronic exposure to toxic agents that induce neuronal degeneration¹² and ischaemia¹³ may lead to an understanding of the mechanisms of ubiquitin accumulation in inclusion bodies and of its significance in neuronal degenerative diseases. □

1. Mori, H. *et al. Science* **235**, 1641–1644 (1987).
2. Cole, G.M. & Timiras, P.S. *Neurosci. Lett.* **79**, 207–212 (1987).
3. Perry, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 3033–3036 (1987).
4. Lennox, G. *et al. Neurosci. Lett.* **94**, 211–217 (1988).
5. Manetto, V. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4501–4505 (1988).
6. Lowe, J. *et al. J. Path.* **155**, 9–15 (1988).
7. Leigh, P.N. *et al. Neurosci. Lett.* **93**, 197–203 (1988).
8. Lowe, J. *et al. Neurosci. Lett.* **94**, 203–210 (1988).
9. Murri, K.G. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3019–3023 (1988).
10. Hershko, A. *J. biol. Chem.* **263**, 15237–15240 (1988).
11. Bond, U. *et al. J. biol. Chem.* **263**, 2384–2388 (1988).
12. Uney, J.B. *et al. FEBS Lett.* **235**, 215–219 (1988).
13. Magnusson, K. & Wieloch, T. *Neurosci. Lett.* **96**, 264–270 (1989).

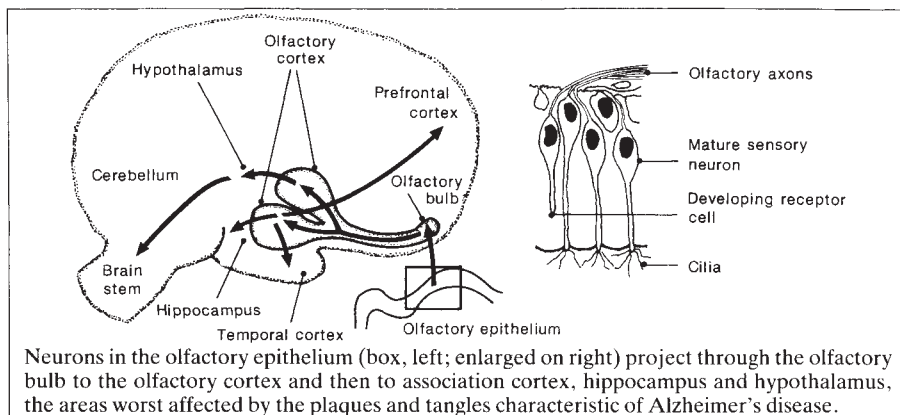
Jean-Marc Gallo is a CNRS staff scientist currently working with Brian Anderton at St. George's Hospital Medical School, London SW170RE, UK.

A nose for Alzheimer's disease?

ALZHEIMER'S disease is one of several types of dementia. Both for research and for clinical management, it is important to distinguish between the various types, but this can be done with certainty only by post-mortem examination of the brain. On page 736 of this issue, B.R. Talamo *et al.* report observations that may help to resolve this dilemma: in people who died of Alzheimer's disease they find pathological changes in the sensory epithelium of the nose. As this tissue is relatively easy to biopsy, such changes may offer a way of diagnosing the

were not, however, seen in olfactory neurons.

The neurons of the olfactory epithelium, which are of central origin, are unusual because they are the only neurons of the mammalian central nervous system known to proliferate throughout adult life (Graziadei, P.P.C. & Monti Graziadei, G.A. *J. Neurocytol.* **8**, 1–18; 1979). This may allow cultures of olfactory neurons to be established in which the molecular changes leading to Alzheimer's disease can be studied directly. The increased phosphorylation of



disease while the patient is still alive.

Some patients with Alzheimer's disease have both an impaired sense of smell and difficulties with olfactory discriminations. This clue led Talamo and her colleagues to examine the olfactory epithelium (see box) with antibodies specific for proteins known to be involved in the neurofibrillary tangles and neuritic plaques that characterize the brains of Alzheimer's patients. They found abnormal masses of neurites (the axonal and dendritic processes of neurons) in the sensory epithelium of eight out of nine patients who had the typical tangles and plaques in the brain. In contrast, neuritic masses were found in only two out of fourteen controls (aged-matched patients dying of other causes and patients with other neurological disorders).

As well as abnormal morphology, antibody staining revealed a change in the composition of subunits of the neurofilament proteins in the axons of the olfactory neurons. In the normal cases, only the dephosphorylated form of the mid-sized subunit was detected, whereas the Alzheimer's cases also had the phosphorylated form of the heavy subunit, a phosphate-independent form of the mid-sized subunit and the light subunit.

The neuritic masses also reacted strongly with the antibodies for the phosphorylated forms and with two antibodies that specifically stain tangles and plaques in Alzheimer's brains (ALZ50 and a monoclonal antibody to tau, a microtubule-associated protein found in the paired helical filaments that form a large part of the neurofibrillary tangles). Typical tangles

neurofilament proteins, for instance, implies changes in protein kinase or phosphatase regulation. The immaturity of olfactory neurons may make them particularly susceptible to factors such as protease inhibitors, some of which promote neurite outgrowth. One form of the precursor protein of the amyloid protein found in plaques contains a serine protease inhibitor-like insert (see the News and Views article by R. W. Carrell in *Nature* **331**, 478–479; 1988) and glia-derived nexin, which is also a serine protease inhibitor, is found in both the peripheral and central parts of the olfactory pathway (Reinhard, E. *et al. Neuron* **1**, 387–394; 1988).

An outstanding question is how early in the disease these changes can be detected. Clearly, longitudinal studies are now needed as well as a thorough examination of patients with other types of dementia. In Alzheimer's disease, the olfactory regions of the cortex, together with the hippocampus and the olfactory parts of the amygdala, are the worst affected by tangles and plaques. R.C.A. Pearson *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **82**, 4531–4534; 1988) have suggested that the disease spreads along the projection neurons from the olfactory cortex to the association areas of the temporal, parietal and frontal lobes (see box); the visual, body sensory and motor parts of the cortex are relatively unaffected. It will be fascinating to see whether the disease originates with a malfunction in the way the growth of olfactory neurons is regulated. Jennifer Altman

Jennifer Altman is an assistant editor of *Nature*.

Astrophysics

Comet showers and γ -ray bursts

Bohdan Paczyński

BRIGHT bursts of γ -rays originating from a single point in the sky continue to puzzle astrophysicists. Typically the position of the sources is known to an accuracy of a few arcminutes at best, making any identification with known objects in the sky impossible. Also, almost all bursters have been observed only once. A few have flashed repeatedly, and on page 716 of this issue¹, Boer, Hameury and Lasota suggest that comets falling onto magnetic, white-dwarf stars could be the cause of these repeaters. Their theory has the unusual merit in this field of predicting testable effects.

There are a few hundred cosmic γ -ray bursts known, and almost a hundred new bursts are discovered every year with many space probes. There is no such thing as a typical burst. The duration of various events ranges from less than 10 milliseconds to more than 1,000 seconds, the typical photon energies may be 30 keV for some, and 3 MeV for others, the limits of all these ranges being instrumental.

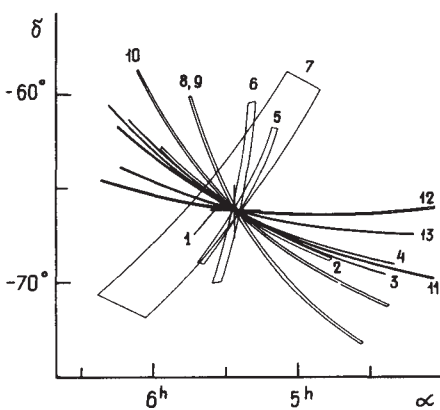
Only three γ -ray bursters are known to repeat: GB790305 (ref. 2), GB790324 (ref. 3) and GB790107 (refs 4,5). The first burst of GB790305 was the famous '5 March 1979' event, by far the most intense of all bursts ever detected, and unique in almost all observable properties. Therefore, even though it was followed by about 15 weaker and softer events over the next 2 years (see figure), it is most likely a different type of object from the other two.

GB790324 was detected on 24, 25 and 27 March 1979, and GB 790107 flashed for the first time on 1 January 1979 and was recorded almost 100 times over the next 5 years. The two differ in the number of repetitions, but otherwise they are remarkably similar: their spectra are unusually soft, and are characterized by a thermal energy of about 40 keV, they are all short, approximately 0.1 s in duration, both were seen close to the galactic plane and both were detected for the first time in the first quarter of 1979. No other soft γ -ray repeater has been discovered since. Some of these remarkable common features are essential, and some are not.

A few explanations for the repeaters have been proposed. For example, I have suggested^{6,7} that the gravitational lensing by a cluster of galaxies of a single non-repeating γ -ray burst could lead to a series of micro-images that arrive at different times in the error box of the repeater. Kouveliotou *et al.*⁸, considering

the directions to the three repeaters, suggest that they could be population I (young, metal-rich) stars in our Galaxy and in the nearby Large Magellanic Cloud. Livio and Taam⁹ have proposed that comets falling onto a neutron star would give the right kind of γ -ray spectrum and repeat frequency. Other models have included starquakes and local thermonuclear runaways on the surface of accreting neutron stars.

The model of Boer *et al.*¹ is similar to that of Livio and Taam⁸. The authors



Error boxes for the positions of 13 bursts from the source of the 5 March 1979 event, detected between 5 March 1979 and 31 December 1982 by detectors on pairs of Venera spacecraft. Pinpointing the source of γ -ray bursts is difficult because of their transient nature. These positions were determined by triangulation. Burst durations varied between 0.1 s and 3.5 s. (From ref. 2.)

propose that the repeaters are due to comets falling onto nearby magnetic white dwarfs. Livio and Taam noted that the spectra of the repeaters are very similar to those of X-ray binary pulsars, which are known to be magnetic neutron stars accreting matter from a binary partner. (The distinction between the 'soft' γ -rays of the bursters and the 'hard' X-rays of the binaries is largely artificial, both occupying a range in the electromagnetic spectrum around 40 keV.) If the accretion was intermittent, as would be the case if comets and not gas were the accreted matter, the source might be a soft γ -ray repeater.

But there is a problem: neutron stars are supernova remnants and any cloud of comets would probably be ejected during a supernova explosion. Boer *et al.*¹ noticed that the spectra of soft γ -ray repeaters are also very similar to those of accreting magnetic white dwarfs observed in binary systems known as polars (AM Her type stars). As the formation of white dwarfs

is a gentle process, there is no reason not to have comets around them. Presumably, on its first passage close to a white dwarf, the nucleus of a comet would be broken up into several pieces, so that on subsequent orbits there would be a series of accretion events, giving rise to recurrent bursts.

The strength, duration, and interval between the bursts are difficult to estimate, and probably impossible to calculate with any precision. The spectrum may also be very difficult to calculate. Fortunately, a semi-empirical approach is possible, and fairly convincing. Observations of polars demonstrate that an accreting magnetic white dwarf produces a hard X-ray spectrum, and the accretion luminosity of polars is known to vary on a very short timescale, so burst-like events should be possible.

If the two soft repeaters are indeed white dwarfs, then they must be very nearby to account for the observed intensity. Boer *et al.* estimate that their distance should be no more than 15 parsecs — a thousand times closer than the Galactic Centre. The faintest (presumably the oldest) white dwarfs known have absolute visual magnitude $M_v \approx 16$ (ref. 10). At a distance of 15 parsecs the faintest white dwarf should have an apparent magnitude of $m_v \approx 17$. The error box for the position of GB790107 is a very elongated ellipse, with a total area of about 0.12 square degrees⁴, and that of GB790324 is only slightly larger³. Therefore, it should be a fairly straightforward matter to look for nearby stars in these areas, identifiable by their high proper motion. The only problem is that the fields are very crowded, right in the Milky Way.

The new theory has a very clear prediction. If two magnetic white dwarfs are discovered in the two small error boxes then they will have to be associated with the two soft γ -ray repeaters, no matter how difficult it might be to work out theoretical details. The *a priori* probability of finding two magnetic white dwarfs in two randomly placed error boxes is only 10^{-9} (ref. 11). If no white dwarfs can be seen, then the theory is dead, no matter how well the details can be worked out. □

- Boer, M., Hameury, J.M. & Lasota, J.P. *Nature* **337**, 716–718 (1989).
- Golenetskii, S.V., Ilyinskii, V.N. & Mazets, E.P. *Nature*, **307**, 41–42 (1984).
- Mazets, E.P., Golenetskii, S.V. & Gur'yan, Yu. A. *Soviet Astr. Lett.* **5**, 343–344 (1980).
- Atteia, J.-L. *et al.* *Astrophys. J.* **320**, L105–L110 (1987).
- Laros, J.G. *et al.* *Astrophys. J.* **320**, L111–L115 (1987).
- Paczynski, B. *Astrophys. J.* **317**, L51–L55 (1987).
- Paczynski, B. *Astrophys. J.* **308**, L43–L47 (1986).
- Kouveliotou, C. *et al.* *Astrophys. J.* **322**, L21–L25, (1987).
- Livio, M. & Taam, R.E. *Nature* **327**, 398–400 (1987).
- Liebert, J., Dahm, C.C. & Monet, D.G. *Astrophys. J.* **332**, 891–909 (1988).
- Angel, J.R.P., Borra, E.F. & Landstreet, J.D. *Astrophys. J. Suppl.* **45**, 457–474 (1981).

Bohdan Paczyński is at Princeton University Observatory, Peyton Hall, Princeton, New Jersey 08544, USA.

Evolution

Responses to chemical warfare

J. S. Jones

SPECIES, like states, often indulge in arms races; and their economy may become so involved in the process that it determines the course of evolution. The struggle between humans and insect pests has such a history. After some initial easy victories, mutual escalation has meant that many crops now demand repeated spraying and that 3 per cent of a pest's proteins may be devoted to detoxification. This race has gone on for only as long as that between the West and the Warsaw Pact. There has been much more prolonged chemical warfare between insects and plants, and an equivalent coevolution of natural pesticides and of means of detoxifying them. Indeed, some insects have evolved methods of dealing with plant toxins which make them resistant to chemical control. At a recent meeting in Australia* — where most organisms seem to have evolved the ability to bite, sting or poison their fellows — the parallels between the genetics of resistance in pests and that of the use of toxic resources by fruitflies were discussed. There is a real prospect of economically valuable results emerging from fundamental research.

Many *Drosophila* grow on poisonous plants. Local adaptation to a toxin may lead to rapid genetic change. *D. mojavensis* lives on agria cactus on the Baja California peninsula, but on organpipe on the Mexican mainland. Flies choose a medium supplemented with a chemical cocktail close to the toxins of their natural host (J.C. Fogelman, Univ. Rochester). There are also differences in the rate of development on each host (W.J. Etges, Univ. Arkansas), and the beginnings of reproductive isolation between the races. Mainland females discriminate against Baja males when given a choice of mates by using chemical cues in the cuticular hydrocarbons (T.A. Markow, Arizona State Univ.). There may be a balance between sexual and natural selection, as lower-molecular-weight hydrocarbons increase male attractiveness but are less effective at retaining moisture. Attractiveness decreases at high temperature, as this increases the proportion of long-chain hydrocarbons. Males with heavyweight waterproofing are impervious but impotent. In *D. melanogaster*, resistance to desiccation is achieved only at the cost of evolving a lower metabolic rate (P.A. Parsons, Griffith Univ.) and the shift to a new toxic host by *D. mojavensis* may have been forced on the flies as the Mexican

climate dried. The process is going on apace in the cactus *Drosophila* of the Caribbean, where it has led to speciation (W.B. Heed, Univ. Arizona).

The interaction between fly and toxin involves some third parties: the bacteria and yeasts which produce the cactus rots upon which the flies feed. Many of these are restricted to a particular cactus, and by detoxifying poisons speed the growth of the *Drosophila* (W.T. Starmer, Syracuse Univ.). In return they are transported from plant to plant by the flies.

Some cacti have themselves become pests. The prickly pear *Opuntia* has been accompanied by *D. buzzatti* in its spread across the world. Although the dependence of the flies on yeast and cactus is just as rigid as in *D. mojavensis*, the interaction between host choice and fly fitness is much more complex (J.S.F. Barker, Univ. New England). In other *Drosophila* that feed on toxic plants such as fungi, the genetics of host specialization is also complicated and may be modified by experience (J. Jaenike, Univ. Rochester). *D. melanogaster* uses a wide range of foods. Many contain ethanol, a toxin to which this species, like its human commensal, is well adapted. The genetics of host preference in *D. melanogaster* has proved subtle and elusive (A.A. Hoffman, La Trobe Univ.). New population genetic theory shows that any tendency of animals of a particular genotype to choose a habitat in which they are relatively fit could be a potent mechanism of maintaining genetic diversity, as long as the same gene influences habitat selection and fitness (P.W. Hedrick, Pennsylvania State Univ.). There is as yet no clear evidence of such pleiotropy in the genes for toxin resistance and for host selection. As is so often the case in population genetics, the theoretical cart is well ahead of the empirical horse.

Much is known of the biochemistry of detoxification of natural and artificial poisons. The alcohol dehydrogenase gene is central to ethanol detoxification, and genotypes which increase flux through this pathway are more common in wineries and survive better on alcohol (B.W. Geer, Knox College.). *D. melanogaster* has one structural gene with two promoters active at different times during development; but in the cactus-fly species the structural gene has been duplicated, and each copy has a promoter homologous to that of *D. melanogaster* larvae (D.T. Sullivan, Syracuse Univ.). There may be a parallel here with the gene amplifications often involved in insecticide resistance.

Detoxifying esterases have also undergone gene duplication in *Drosophila*. The duplicated esterases within cactus *Drosophila* species show more sequence similarity than does each of them among species, suggesting that concerted evolution is involved (P.D. East, Univ. New England). The *Est-6* locus in *D. melanogaster* has more than 20 alleles differing in amino-acid sequence. The commonest allele has the same DNA sequence throughout the world, although other alleles show sequence divergence. This common allele may represent a new esterase which has spread so rapidly that there has been no time for sequence change (J.G. Oakeshott, CSIRO, Canberra). Although the nature of selection on this allele is unknown, sexual selection is a possibility as esterases secreted by males act as molecular contraceptives which inhibit female remating (R.C. Richmond, Indiana Univ.).

Modern genetic technology means that cloned genes in one species can be used to search for their homologues in others. The evolution of *Drosophila* in response to natural toxins may therefore have practical use in studying insecticide resistance. The sheep blowfly *Lucilia coprina* in Australia is resistant to organophosphates because of the evolution of new esterases. At least two loci are involved and, as in *Drosophila*, they are part of a gene family. The chromosomal positions of these genes and their activity during development are the same as in *Drosophila*. A search for the *Lucilia* resistance genes using a *Drosophila* esterase probe is the first step in cloning these genes and perhaps in developing new methods of combating their effects (R. Russell, CSIRO, Canberra). *Lucilia* eye-colour genes also have amino-acid sequences similar to those of their equivalents in *D. melanogaster*. But this homology conceals considerable divergence in DNA sequence. *Lucilia* and *Drosophila* have very different codon preferences, and the transcription unit for the eye mutants in *Lucilia* is several times longer than in the more widely studied species because of the presence of much longer introns (A. J. Havells, Australian National Univ.).

More than 500 insect pests are now resistant to chemical control and there is even evidence of a recent spread of resistance through the world's *D. melanogaster*. Most of the genes involved are single mutants of large effect, which are just the sorts of genes which can now be genetically manipulated in *Drosophila*. It is inconceivable that new weapons in the war on pests will come from an understanding of insects' own defences against natural chemicals. Perhaps this arms race will be the last in human history to end in the destruction of only one of the antagonists. □

J.S. Jones is in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK.

* US-Australia Conference: Ecological and Evolutionary Genetics of *Drosophila*. University of New England, Armidale, New South Wales, Australia. 3-10 January 1989.

Cloud physics

Growth of large hailstones

Anthony Illingworth

HAIL causes considerable damage to agriculture: for example, in France losses are estimated at £100 million per annum and in Italy the damage in the Po valley alone is even greater. Although hail storms originate in cumulonimbus clouds, not all large thunderstorms produce hail, and a cloud has to perform a delicate

balancing act if it is to grow hailstones more than 1 centimetre in size. Many people have been tempted to try to upset this balance and suppress the hail by firing cannon or rockets into the clouds. But our knowledge of hail growth is incomplete and because we cannot precisely predict what would have happened to the storm without any interference, these hail-control techniques are untested. L. Cheng and D.C. Rogers (*J. atmos. Sci.* **45**, 3533–3545; 1988) report observations of radar data, time-resolved surface collections of hail and cloud photographs of a storm in Alberta, Canada, which confirm earlier suggestions that the growth of large hail is a two-stage process.

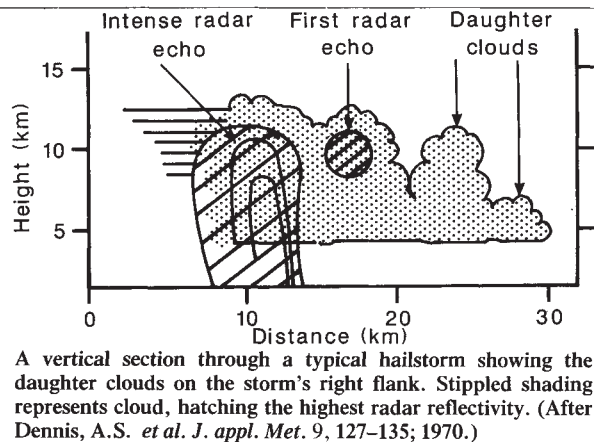
The main problem is to explain how the embryo of a hailstone can be confined in a region dense with liquid cloud droplets for sufficient time to grow to a hailstone. A falling hailstone grows by sweeping out these droplets which freeze as soon as they hit the hailstone. The maximum amount of cloud water produced by cooling and condensation of moist, warm surface air as it is lifted to -20°C is about 4 g m^{-3} , and a hailstone needs to spend about 20 min in such a cloud if it is to grow to 1–2 cm in size.

Such hailstones have a terminal velocity of about 20 m s^{-1} or so, but updraught velocities of this magnitude occur in thunderstorms, and so the final stage of growth to the larger size can occur as the hailstone is balanced against the updraught within the cloud and then descends to the ground. The difficulty arises in the early stages of growth. It takes several minutes to achieve a size of a few millimetres, and during this time the embryonic hailstone will have a low terminal velocity, and if it finds itself in a 20-m s^{-1} updraught, it will pass straight through the cloud.

In Alberta, three vehicles chased the storm and collected hailstones underneath the regions of high radar reflectivity. At any one place the large fall lasted only a few minutes. The maxima in the radar reflectivity of the cloud could be traced backwards in time for 20 minutes to the new growth regions of the storm 30 km

away from the main hail-producing updraught. Airborne photography confirms the existence of a series of these small but rapidly growing cumulus clouds, called daughter or feeder clouds, in the form of a staircase (see figure).

The picture emerges, then, of hail embryos which can be in the form of



raindrops or small ice pellets, growing to about 5 mm while suspended in the weak updraughts of the growing daughter cells. The daughter cell then grows to become the principal updraught, or merges with the main storm, and the final growth stage occurs in a region of much higher vertical air velocity.

Problems do remain. First, why does the succession of daughter cells form with

a separation of about 3 km? Cheng and Rogers note that the line of daughter cells was oriented along a line approximately parallel to the environmental wind shear, consistent with the resonance-mode theory of shear-induced instability proposed by D.P. Lalas and F. Einaudi (*J. atmos. Sci.* **33**, 1248–1259; 1976).

Second, why are so few embryos introduced into the main updraught? The concentration of raindrops in mature clouds is several thousand per cubic metre, and if they all found their way into the main updraught they would collect cloud droplets so quickly that the supply of cloud water in the updraught would soon be exhausted. Instead of a few large hailstones, many more smaller hail pellets would be produced.

My recent polarization radar measurements of developing cumulus clouds (*Nature* **336**, 754–756; 1988) suggest that the first detectable radar echo consists of just a few large raindrops in concentrations of less than one per cubic metre. If this is so, then these raindrops would form natural embryos for large hailstones without invoking any elaborate mechanism for size sorting. But we still cannot predict whether any attempt to introduce artificial nuclei such as silver iodide into daughter cells would enhance or diminish the hail risk, or whether such efforts at modification are futile. □

Anthony Illingworth is in the Department of Physics, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK.

Oceanography

Are rising and falling particles microbial elevators?

J.R. Toggweiler

A CASUAL survey of sediments from the deep-sea floor reveals that most of the sedimentary material consists of the remains of tiny plants and animals that once lived near the surface. Not long ago it was widely assumed in oceanography that particulate material sinking from the 'euphotic' upper layers of the ocean was composed of fairly indigestible organic particles and mineral precipitates, like CaCO_3 , which slowly settled through a dark and largely lifeless abyss. Since sediment traps were developed in the late 1970s to catch the sinking material in mid-water, it has become apparent that material sinking through the abyss is composed not of inert, slow-settling particles, but rather of fast-sinking,

nutritious material which can feed a rich assortment of deep-dwelling animals, both in the water column and at the sea floor. Now we learn that the flow of particulate organic material goes both ways: K.L. Smith Jr, P.M. Williams and E.R.M. Druffel report elsewhere in this issue (*Nature* **337**, 724–726; 1989) that the deep-dwellers of the abyss are generating buoyant organic particles which ascend toward the surface. Their data show that the upward flux may be 15–20 per cent of the sinking flux on average.

Smith *et al.* hung moored sediment traps, both in the usual upward-facing mode and upside down, at 600 m and 1,600 m above the sea floor at two locations in the Pacific Ocean. Besides finding

a surprisingly large amount of material in the inverted traps, they found that ascending material is distinctly different from the sinking material at the same depth. The ascending material consists of egg masses and amorphous gelatinous material, and includes none of the skeletal tests and clay particles common in the sinking flux (K. Smith, personal communication). Furthermore, the upward flux seems to have a distinct seasonal signal, with larger fluxes observed in the spring at both locations.

The sinking flux of organic matter 3,000 m or so below the surface is a small fraction of sinking fluxes in the upper kilometre (Martin, J. H. *et al.* *Deep-Sea Res.* **34**, 267–285; 1988). What remains to be determined is whether the 0.15–0.20 ratio of upward to downward transport identified by Smith *et al.* is representative of the upper kilometre. Should this relationship hold where downward fluxes are high, then the upward flux would have to be considered a significant component of global carbon and nutrient budgets. Smith *et al.* did hang a set of normal and inverted traps just 500 m below the surface, but strong upper-ocean currents bent the mooring line so severely that the catches in both traps were completely compromised.

The composition of the ascending material in the traps of Smith *et al.* and its apparent seasonal cycle suggest very strongly that the upward flux forms part of the life cycles of certain deep-sea animals. We do not know whether the ascending egg masses caught 3,000 m or more from the surface actually travel all the way up and develop into planktonic juvenile forms. The eggs could stop rising at some depth and develop in mid-water.

In any event, the adult forms must eventually descend to the near-bottom environment to complete the cycle. M.W. Silver *et al.* (*Nature* **309**, 246–248; 1984) discovered large populations of ciliated protozoa associated with sinking detritus in the upper 2,000 m of the Pacific. They found distinct species changes among these populations at different depths. Silver (personal communication) believes that the protozoa locate and colonize detrital particles in mid-water and feed as they ride the particles down. Rather than ride completely out of their habitats they release themselves and ascend back to their original depth range. The behavioural strategies which are suggested by the studies of Smith *et al.* and Silver *et al.* define a conceptual model for the transport of organic matter which makes the upward flux seem not so mysterious after all. □

J.R. Toggweiler is in the Geophysical Fluid Dynamics Laboratory, Princeton University, PO Box 308, Princeton, New Jersey 08542, USA.

Physiological ecology

Species borders and metabolism

Jared M. Diamond

WHAT limits the distribution and abundance of species? Different answers have been suggested: food abundance, habitat structure, interspecific interactions such as competition and predation, and physiological limitations imposed by abiotic factors such as temperature. Temperature is one putative controlling variable that cannot be manipulated experimentally at large outdoor study sites. Thus, Terry Root (*Ecology* **69**, 330–339; 1988) has

metabolic rate are available as a function of ambient temperature for winter-acclimatized individuals of 14 species of songbirds whose northern border thus coincides with an isotherm. Above a certain critical temperature characteristic of each species, resting metabolic rates of birds remain at the basal metabolic rate and are independent of ambient temperature. Below this critical temperature, resting metabolic rate increases linearly with

decreasing temperature to supply the extra energy needed to balance heat loss. Root examines the resting metabolic rate measured for each species at the minimum January temperature of its northern boundary. Ratios of this rate to the basal metabolic rate of the same species vary tightly around 2.49 with a standard error of only 0.07. She obtains virtually the same ratio for 36 other songbird species whose metabolic rates, though not measured directly,

can be estimated from the dependence of metabolic rate on body weight.

Thus, these 50 North American songbird species are limited to wintering in areas where they do not have to raise their resting metabolic rate beyond 2.5 times the basal level in order to stay warm — the '2.5 rule'. This strikingly consistent value applies to songbirds from wrens to crows, ranging in weight from 5 to 448 grams and in diet from seed-eaters to insectivores, and with winter northern borders ranging from Florida to Canada. Now recall G.E. Hutchinson's definition (*Cold Spring Harb. Symp. quant. Biol.* **22**, 415–427; 1957) of the fundamental niche as the set of environmental conditions under which a species would be physiologically capable of maintaining itself, if it were not for restrictions imposed by interspecific interactions such as competition and predation. Root's discovery in effect suggests an operational definition of the fundamental niche for temperature-limited songbirds. It thereby allows several questions to be formulated in terms of physiological ecology.

First, what is the physiological basis of the 2.5 rule? Does it imply that an organism can synthesize only enough extra intestinal nutrient transporters or enzymes of intermediate metabolism to process nutrients at a rate 2.5 times the resting rate?

Second, does a similar rule apply to



The northern boundary of distribution around Christmas for the Eastern Phoebe (solid line) lies very close to the -4°C isotherm of January minimum temperature (dashed line).

made an important advance by her discovery of a temperature-dependence of some species borders that is strikingly consistent and closely tied to physiological mechanisms.

Root has analysed the Christmas bird counts organized annually since 1900 by the National Audubon Society. (Each year on a day within 2 weeks of 25 December, at 1,317 sites throughout the United States and Canada, hordes of fanatical bird-watchers count all birds observed within a 15-mile-diameter circle.) For each species and site for the decade 1962–63 to 1971–72, Root (*J. Biogeogr.* **15**, 489–505; 1988) calculated the average number of bird individuals seen per hour per group of bird-watchers, and plotted this measure of abundance on contour maps for comparison with contour maps of six environmental variables.

One of these variables is winter temperature, taken as average minimum temperature at the site during January. Root finds that more than half of the bird species analysed (60 per cent) have the entire length of their northern range border during Christmas counts closely associated with a particular isotherm. The Eastern Phoebe, for example, winters in the south-eastern United States north-west to a limit coinciding closely with the -4°C minimum January isotherm (see figure).

Published measurements of resting

upper limits of altitude for those breeding montane bird species that are restricted by temperature? This question could easily be tested by examining summer minimum temperatures at birds' altitudinal ceilings and by comparing the polewards decrease in ceiling with that in temperature.

Third, might there be a ceiling similar to the 2.5 rule for increased metabolic rates in response to variables other than low temperature? Winter is not necessarily the period of highest energy expenditure for a bird; expenditure also peaks when parent birds are feeding young and perhaps during moulting. It is intriguing, as Root notes, that avian energy expenditures during reproduction and those averaged over the whole year also seem to be low multiples of the basal metabolic rate (up to 4 times and 2.6–3.2 times, respectively).

Of course, Root does not claim that all range borders are set by temperature. She shows that some coincide closely with isotherms, some with contours of other environmental variables, and others do not coincide with any environmental contour. J.W. Terborgh (*Ecology* **66**, 1237–1246; 1985), studying the world's most

species-rich avifauna, that of the equatorial Andes, showed that about two-thirds of bird altitudinal limits there are set by interspecific competition, and less than one-sixth are set directly by physical gradients like that of temperature. Terborgh's finding supports Hutchinson's distinction between the realized niche of a species (that compressed by competition and predation) and its fundamental niche (that permitted by its physiology). The more species-rich the fauna, the narrower on the average will be the realized niche compared with the fundamental niche.

The limits of the fundamental niche are not achieved in practice by most Andean bird species, and the same may also be true for some temperate-zone congeneric species pairs with abutting ranges (for example, various pairs of chickadees and jays in North America). Will the ratio of resting metabolic rate at northern winter boundaries or altitudinal ceilings to basal metabolic rate in these competition-limited species be much less than 2.5? □

Jared M. Diamond is a professor of Physiology at the University of California Medical School, Los Angeles, California 90024-1751, USA.

Anthropology

Early teething troubles

Paul G. Bahn

RECENT studies of grooves in ancient human teeth have led to a widespread consensus that the toothpick may be one of the earliest tools devised by humanity. Although this idea is not new, it is controversial. Similar grooves had been found and reported¹ in 1911 at the French Middle Palaeolithic (200,000–35,000 years before present (BP); Neanderthal period) site of La Quina, and interpreted as the result of constant use of toothpicks. But others have attributed the grooves to chemical erosion. The new studies^{2–4} strongly suggest that picking one's teeth is indeed one of humanity's oldest habits.

The grooves generally occur along the junction between the cementum and the enamel, mostly on premolars and molars (both upper and lower), and have a semi-circular or trough-like shape, with longitudinal and parallel striations. In some cases the grooves are bordered by ridges of reactive cementum, probably produced as a response to the irritation caused by the tool. The earliest known examples of the phenomenon occur in upper premolars of *Homo habilis* at Omo (Ethiopia), dating to 1.84 million years ago⁵. Several cases in *H. erectus* and archaic *H. sapiens* are known from the Middle Pleistocene (700,000–130,000 years BP) at sites such as Zhoukoudian (China), Rabat (North Africa), Atapuerca (Spain) and in the Soviet Union^{2–4}.

The most thorough investigation has been carried out by Frayer and Russell⁶ on Neanderthal teeth from Krapina, Yugoslavia, which date to the Middle Palaeolithic. Their microscopic examination of shallow grooves on 14 teeth from 10 individuals (aged between 16 and 27) reveals that they existed only on erupted permanent teeth, almost exclusively on premolars and molars, and most commonly on third molars. The same phenomenon has also been reported on isolated teeth from other Middle Palaeolithic sites: not only La Quina¹ but also Hortus (France) and Gibraltar^{2,4}. Other European Upper Palaeolithic (35,000–130,000 years BP) cases are known, particularly in the

skeletal material from the Grimaldi caves (Italy)⁶, where the grooves are more common in upper teeth than in lower and, as usual, in back teeth rather than incisors.

The grooves persist in Europe through the Mesolithic, Neolithic and Bronze Age, and into history. In the New World too, they have been found as far back as the 7th millennium BP in North America, and in the 10th millennium in Peru⁷.

There are alternative explanations for the phenomenon — root caries (dental erosion, where bacteria attack the tooth surface); gritty saliva being propelled between the teeth; and fibre processing. None of these, however, really fits the evidence. Taphonomic erosion would not cause ridges of reactive cementum, root caries would not extend onto the enamel surface, and gritty saliva would probably affect the front teeth as well, while fibre processing might affect them even more than the back ones. The consistent location of the grooves; the longitudinal polish and striations, suggesting a prolonged back-and-forth movement of an inflexible probe; and the similarity of the prehistoric grooves to those documented in historical and modern populations including Amerindians, Australian Aborigines, Canary Islanders and Upper Dynastic Egyptians, are all factors that argue for the toothpick interpretation of the data.

Why were toothpicks used by early humans? Oral hygiene is known even among some non-human primates⁷; but it is thought that the presence and depth of some prehistoric grooves reveal a prolonged repetitive activity. Some ascribe this to attempts to relieve problems such as gum irritation, but others⁶ speculate that it may have been a largely unnecessary, non-functional pastime that was cultural rather than practical. □

1. Siffre, A. *Bull. Soc. Préhist. fr.* **8**, 741 (1911).
2. Puech, P.-F. & Cianfarani, F. *Curr. Anth.* **29**, 665 (1988).
3. Eckhardt, R.B. & Piermarina, A.L. *Curr. Anth.* **29**, 668 (1988).
4. Turner, C.G. *Curr. Anth.* **29**, 664 (1988).
5. Frayer, D.W. & Russell, M.D. *Am. J. phys. Anth.* **74**, 393 (1987).
6. Formicola, V. *Curr. Anth.* **29**, 663 (1988).
7. McGrew, W.C. & Tutin, C.E.G. *Nature* **241**, 477 (1973).

Paul G. Bahn is a freelance writer on Archaeology at 428 Anlaby Road, Hull HU3 6QP, UK.

100 years ago

A MOVABLE ZOOLOGICAL STATION

IN Bohemia, much attention has been given for more than twenty years to the study of the fauna of ponds and lakes, but the work has been rendered difficult by the impossibility of the organisms being examined instantly in their habitats. Last year, a little movable station, suitable for real biological work was presented to the Committee for the Physical Exploration of Bohemia; and there is good reason to hope that the use of this structure may be attended by important scientific results.

From *Nature* **39**, 416; 28 Feb. 1889.



Photosynthesis

Casting new light on old relationships

Peter D. Moore

THE ecological significance of the biochemical routes by which photosynthetic plants fix atmospheric carbon has been the subject of considerable debate. It now seems that the simple classification of plants into one of three photosynthetic types, C3, C4 and CAM, is less distinct than has been thought. Recent experiments by Ueno *et al.*¹ and by Groenhof *et al.*² suggest that closely related taxa may differ in their photosynthetic pathways, and that in some plants C4- and CAM-type systems can be induced by appropriate environmental stresses.

Under conditions of high light, high temperature and drought, plants that use the enzyme phosphoenolpyruvate carboxylase (PEPC) for the initial fixation of carbon — C4 and CAM plants — are at a distinct advantage over C3 species. PEPC allows CAM plants to fix carbon in the dark, enabling them to keep their stomata closed during the day. In C4 plants, there is a spatial rather than a temporal separation of the initial and final fixation process via the Calvin cycle, which takes place only in specialized vascular bundle sheath cells, an arrangement termed Kranz anatomy. Both types of photosynthesis have the advantage of the high affinity of the PEPC enzyme for carbon dioxide, and thus have low carbon dioxide compensation points. They also avoid the problems of photorespiration experienced by C3 plants. The ability to scavenge for carbon results in a further potential ecological advantage of PEPC-based systems — they can compete efficiently in aquatic habitats where carbon is often in short supply.

It has long been recognized that closely related taxa may differ in their photosynthetic systems. In 1972 Hatch *et al.*³ described hybridization experiments between *Atriplex patula* ssp. *hastata* (a C3 species) and *Atriplex rosea* (a C4 species). They studied isoenzyme patterns of hybrids and concluded that "at least some of the enzymes necessary for the operation of the C4 pathway are unique isoenzyme entities evolved specifically for that purpose".

Frean and Marks⁴ have examined taxa that are even more closely related, indeed that are considered mere variants of a single grass species, *Alloteropsis semialata*, and yet which exhibit C3 and C4 photosynthesis and have normal and Kranz anatomy, respectively. Chromosome squashes of pollen grains from these two variants, however, show that the C3 form

is a fertile diploid ($2n = 18$) and the C4 form is an allohexaploid ($2n = 54$). So the taxonomic difference is greater than one might have supposed and involves polyploidy. Perhaps they should even be regarded as separate species.

Groenhof, Bryant and Etherington⁵ have brought attention to a succulent CAM species, *Sedum telephium*, which can move from a photosynthetic mixed economy with both day and night carbon fixation to a full nocturnal CAM mode when subjected to water stress. They now show² that this switch is accompanied by a substantial increase in dark-period PEPC activity. They also show that two species of the PEPC protein are involved, one (the CAM species) a monomer and the other a dimer.

Progress is also being made on the aquatic front. Ueno *et al.*¹ have described a species of spike rush (*Eleocharis vivipara*) from Florida which can grow in both terrestrial and submerged conditions. They find that the terrestrial form has specialized sheath cells and conducts C4 photosynthesis. But the submerged form appears to fix carbon using the Rubisco enzyme and has anatomy more characteristic of a C3 species. They carried out reciprocal experiments in which they put submerged forms in terrestrial conditions and vice versa, and the plants reverse both their anatomical and biochemical forms.

These results contrast with the findings of Keeley and Busch⁶ using the pteridophyte *Isoetes*, where the submerged form uses a CAM system of photosynthesis. Because the great advantage of carbon fixation by PEPC in aquatic conditions is the ability to compete for carbon when it is in short supply, often in the middle of the day, it would be interesting to know whether *Eleocharis vivipara* retains a C3 system in submerged conditions when starved of carbon. The results do suggest, however, that not only is the biochemical pathway of C4 inducible, but so is the Kranz anatomy. □

1. Ueno, O., Samejima, M., Muto, S. & Miyachi, S. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6733–6737 (1988).
2. Groenhof, A.C., Bryant, J.A. & Etherington, J.R. *Ann. Bot.* **62**, 187–192 (1988).
3. Hatch, M.D., Osmond, C.B., Troughton, J.H. & Bjorkman, O. *Yb. Carnegie Instn. Wash.* **71**, 135–141 (1972).
4. Frean, M.L. & Marks, E. *Bot. J. Linn. Soc.* **97**, 255–259 (1988).
5. Groenhof, A.C., Bryant, J.A. & Etherington, J.R. *Ann. Bot.* **57**, 689–695 (1986).
6. Keeley, J.E. & Busch, G. *Plant Physiol.* **76**, 525–530 (1984).

Peter D. Moore is Reader in Ecology in the Division of Biosphere Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

Daedalus

Strains of sweet music

WHY do we like music? Alone among aesthetic experiences it means nothing, and has no obvious grounding in our instinctive nature. And yet it is probably the most moving and magical of all the arts.

Daedalus reckons that our musical sense evolved when we still lived in trees. The simplest way of telling if a tree is rotten, say, or a branch strong enough to swing on, is to sense its vibrations. A stable structure, well inside its elastic limit, will vibrate linearly, with harmonics in integral ratios and no intermodulation. A branch on the point of breaking, however, will be grossly nonlinear. Its discordant harmonics, and its intermodulation of superimposed frequencies, must bring immediate instinctive alarm to a tree-creature venturing onto it. The fundamental vibrations of a tree can only be sensed with hands or feet, but our ancestors must soon have learnt to listen for higher modes, or to tap or strike the wood to excite them. A tuneful harmonic outcome spelt safety; a discordant one deadly peril. And so the love of harmony entered the human soul. Some people can still appreciate music purely by its vibrations against the skin, and we all instinctively judge the soundness of an object by giving it an exploratory bash.

So, says Daedalus, creatures who live in trees should be musical. Song-birds are obviously on his side; and DREADCO's zoologists are presenting a course of musical appreciation to an audience of koala bears, three-toed sloths, squirrels, tree-frogs, and so on, to look for interest in these species as well. But his theory has clear technological implications too. It suggests a simple, non-destructive means of stress-testing any structure *in situ*. Just vibrate it at a specific frequency, and listen to the harmonics.

A trained ear should be good enough. The old violin makers used to tap a tree to judge if its wood was fit for their purpose. Those hi-fi fans who are really offended by 0.01 per cent of intermodulation distortion should easily be able to sense how far into its nonlinear region a piece of wood has been stressed. But electronic frequency measurement is more accurate still. DREADCO's technicians are testing a simple device that injects two tones into a structure at any point, and measures their intermodulation product: a measure of the local curvature of the stress-strain curve. This shows where the material is on that curve, and hence — what the engineer really wants to know — how much more stress it can take. Scanned over bridges, vehicles and all sorts of similar structures in service, the DREADCO anharmonograph will detect the highly stressed bits before they break.

David Jones

Transatlantic spread of seal virus

SIR—Kennedy *et al.*¹ refer to the “annual migration of porpoises from the Baltic to the North Sea and south to the English Channel” as a possible means whereby the current seal epizootic was spread, and say that “porpoises are known to cross the Atlantic Ocean, so marine mammals along the American continent could be at risk”. They suggest that their finding of morbillivirus infection in two recently stranded harbour porpoises “may explain the declines in porpoise and dolphin populations in European waters in recent years”.

These remarks are surprising because although migration was originally suggested as a possible explanation for the pattern of stranding records², more recent work indicates that UK harbour porpoise strandings mainly reflect the washing ashore of animals accidentally killed during fishing operations³. Porpoises are known to be coastal and fairly sedentary animals. Morphometric and meristic evidence indicates separate eastern and western North Atlantic populations, and Dutch, Baltic and UK North Sea animals may form distinct sub-populations⁴. It therefore seems most unlikely that individual porpoises travel long distances, or cross the Atlantic Ocean.

Kayes⁵ and I⁶ have independently analysed the available sighting and stranding data for the southern North Sea from 1913, and conclude that they can neither support nor exclude the anecdotal evidence of decline in the porpoise and dolphin populations. The records of stranded porpoises have not shown the increase expected had debilitating viral infection long been present in the species. It seems that we need to look elsewhere for the vector of the current seal epizootic.

M. KLINOWSKA

Research Group in Mammalian
Ecology and Reproduction,
Physiological Laboratory,
Cambridge CB2 3EG, UK

KENNEDY *ET AL.* REPLY—Our suggestion of a possible role for porpoises in spread of the seal virus¹ was only partly based on claims of migration and transatlantic crossing by porpoises^{2,6}. We have no expertise in cetacean behaviour and thus accept Klinowska's contention that her own studies probably invalidate these earlier claims. However, morbilliviruses are highly infectious and one infected animal may be sufficient to infect a susceptible population. Furthermore, little is known about the movements of individual porpoises. So, regardless of specific migration patterns, porpoises must be considered as possible vectors of the virus for seals and other marine mammals. This does not imply that they are the only potential vectors.

Evidently, not all biologists agree that porpoises are “fairly sedentary animals” and that it is “unlikely that individual porpoises travel long distances”, since Kayes⁵ states that “they travel over considerable distances” and the possibility of Icelandic porpoises coming into contact with Irish and British porpoises has been mentioned⁴. Evidence⁵ indicating that large-scale intermixing of eastern and western North Atlantic porpoise populations does not occur, does not preclude interchange of small numbers of porpoises enough for transatlantic spread of the virus.

Klinowska's interpretation of Kayes's critical review⁵ of anecdotal evidence is surprising since Kayes states “the only reasonable conclusion that can be drawn from the available evidence is that both the harbour porpoise and bottlenosed dolphin had become rare in the southern North Sea by 1970”. He also concludes that they have “become rare on the Atlantic coast of France, Spain and Portugal” and have “virtually disappeared from the Baltic Sea”.

We do not know how long morbillivirus infection has been present in porpoises. If the virus has been present for a long time, one would expect many porpoises now to be immune and hence high present-day mortality would be unlikely. As detailed necropsies have been done on relatively few stranded porpoises, including those

attributed to fishing-induced injury, the incidence of morbillivirus infection in this species is unknown. On the other hand, recent introduction of the virus to porpoises may be expected to produce high mortality, although not on the scale of the recent seal epizootic, as opportunities for lateral spread of the virus in porpoises are probably fewer than in seals. We have diagnosed morbillivirus infection and found distemper-like lesions in six porpoises from the Irish Sea in recent months. This suggests recent introduction of the virus to these porpoises. It will be of interest to know how the 1988 European stranding records compare with former years.

S. KENNEDY

J. A. SMYTH

P. F. CUSH

S. J. MCCULLOUGH

G. M. ALLAN

Veterinary Research Laboratories,
Stormont, Belfast BT4 3SD, UK

S. MCQUAID

Department of Pathology,
Queen's University,
Belfast BT7 1NN, UK

1. Kennedy, S. *et al.* *Nature* **336**, 21 (1988).

2. Harmer, S. *Rep. Br. Mus. (Nat. Hist.)* **10**, 23 (1927).

3. Klinowska, M. *Rep. int. whaling Comm.* **35**, 45–67 (1985); in *Reasons for Concern. Proc. 2nd North Sea Seminar Vol. 2* (ed. Peet, G.) 73–79 (Werkgroep Noordzee, Amsterdam, 1987).

4. Yurik, D.B. & Gaskin, D.E. *Ophelia* **27**, 53–75 (1987).

5. Kayes, R.J. *The Decline of Porpoises and Dolphins in the Southern North Sea* (Political Ecology Group Ltd, Oxford, 1985).

6. Matthews, L.H. *The Natural History of the Whale* 131 (Weidenfeld and Nicholson, London, 1978).

Preferences of Palaeozoic predators

SIR—We have examined the scars on Palaeozoic trilobites and report that those scars that we attribute to sublethal predation rather than undefined causes are significantly more frequently found on the right side of the trilobites, suggesting that predators preferred to attack that side.

The scars of healed injuries to trilobites

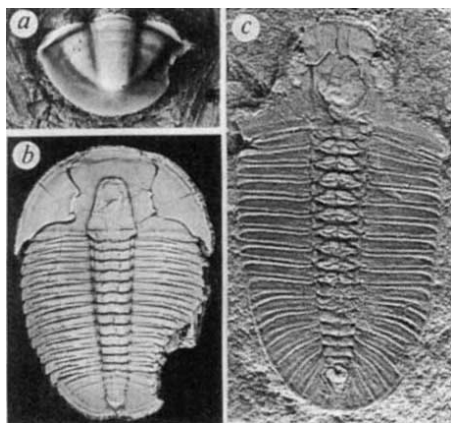
are recognizable as broken areas of the exoskeleton that have become callused. When the scars are in the form of callused embayments on marginal areas (see figure) we attribute them to sublethal predation as they are unlikely to have been accidental breaks.

The areas most susceptible to accidental

Incidence of healed injuries on right, left or both sides of 158 trilobites

	Right side only <i>n</i> (%)	Left side only <i>n</i> (%)	Both sides <i>n</i> (%)
Sublethal predation scars:			
Cambrian trilobites	35 (73)	11 (23)	2 (4)
Post-Cambrian trilobites	21 (64)	11 (33)	1 (3)
All trilobites	56 (69)	22 (27)	3 (4)
Injuries of uncertain origin:			
Cambrian trilobites	15 (56)	10 (37)	2 (7)
Post-Cambrian trilobites	25 (50)	23 (46)	2 (4)
All trilobites	40 (52)	33 (43)	4 (5)
All injuries	96 (61)	55 (35)	7 (4)

Injuries were analysed using a binomial test in which the expected distribution on the left and right sides is 1:1. We analysed specimens showing injuries on one side (78 of 81), and counted specimens showing multiple injuries on the same side only once. Observed right-left ratios of predation scars are 3.2:1 for Cambrian trilobites ($P < 0.001$, $n = 46$); 1.9:1 for post-Cambrian trilobites ($P < 0.055$, $n = 32$); and 2.6:1 for Cambrian and post-Cambrian trilobites ($P < 0.001$, $n = 78$). Injuries of uncertain origin are not statistically significant ($P > 0.100$) for Cambrian ($n = 25$), post-Cambrian ($n = 48$) or pooled ($n = 73$) trilobites. Specimens not in the University of Kansas collection are described in refs 6, 7.



Middle Cambrian trilobites from Utah showing healed arcuate scars resulting from injury during predation, all in collections of the University of Kansas Museum of Invertebrate Paleontology (KUMIP). *a*, *Asaphiscus wheeleri* Meek, pygidium, KUMIP 203889, $\times 2.0$. *b*, *Elrathia kingii* (Meek) with a regenerated, abnormally elongated, pleural tip, KUMIP 204773, $\times 0.9$. *c*, *Bathyriscus elegans* (Walcott), moulting exoskeleton lacking free cheeks, arrow indicates scar, KUMIP 204774, $\times 0.7$.

damage include spines, facial sutures, bilamellar fringes and tips of thoracic segments, so we treated scars in these areas as of uncertain origin. Among 158 specimens with healed injuries, 81 show scars attributable to predation (see table).

Most marginal predation scars are arcuate in outline (see figure). One widespread Cambrian animal, *Anomalocaris*, has been suggested¹ as a possible predator of Cambrian trilobites, and the arcuate injuries preserved on some trilobites (*a*, *b*

in the figure) are similar in size and shape to the opening between the mouth plates of this enigmatic organism. Trilobite remains have been observed in the gut contents of other Cambrian animals^{2,3}. The fossil record of post-Cambrian marine predators is more diverse and those predators were a continuing source of injury to trilobites until the Permian, when trilobites became extinct.

Sublethal predation scars are more common on the right side of the trilobite body than on the left (see table), whereas locations of injuries of uncertain origin do not significantly differ from a random distribution. Our results indicate that lateralized behaviour had developed by Cambrian time, 500 million years earlier than previously observed. This is the first evidence of behavioural asymmetry in the fossil record, except for hominids, and supports the hypothesis that brain laterality has an evolutionary history^{4,5}.

LOREN E. BABCOCK

RICHARD A. ROBISON

Department of Geology,
University of Kansas,
Lawrence, Kansas 66045, USA

1. Whittington, H.B. & Briggs, D.E.G. *Phil. Trans. R. Soc. B* **309**, 569–609 (1985).
2. Bruton, D.L. *Phil. Trans. R. Soc. B* **295**, 619–653 (1981).
3. Conway Morris, S. *et al. Univ. Kansas paleont. Contr.* **122**, (1988).
4. Dennenberg, V.H. in *Laterality and Psychopathology* (eds Flor-Henry, P. & Gruzelier, J.) 29–38 (Elsevier, Amsterdam, 1983).
5. Bryden, M.P. *Laterality: Functional Asymmetry in the Intact Brain* (Academic, New York, 1982).
6. Owen, A.W. *Trans. R. Soc. Edinb.* **76**, 255–272 (1985).
7. Conway Morris, S. & Jenkins, R.J.F. *Alcheringa* **9**, 167 (1985).

Mating success in male pheasants

SIR—Von Schantz *et al.*¹ report that in a population of Swedish pheasants, male spur length is positively correlated with male viability and with male mating success. It is not clear why males with longer spurs survive better, but their greater mating success appears to be explained in this population by female sexual preference for males with longer spurs, a result supported by experimental shortening and elongation of the spurs.

In an accompanying News and Views², Kirkpatrick comments that this is a surprising and mysterious result. Surprising, because the standard interpretation of darwinian sexual selection predicts that attractive secondary sexual characters, like spurs, that enhance mating success, should decrease male survival³. Mysterious, because if spurs are so beneficial, why are they not very rapidly increasing in males and why are they absent from females?

Kirkpatrick offers two solutions to this puzzle. First, that the measurement of selection by von Schantz *et al.* is unrepresentative and in normal years viability selection acts against long spurs; and second, that there is no additive genetic variance for male or female spur length in this population. As Kirkpatrick states, neither

of these explanations seems likely. I would like to suggest a much more straightforward explanation, that spur length is a condition-dependent character^{4–7}.

Imagine that males vary in quality for environmental or genetic reasons and that, other things being equal, higher-quality individuals survive better. High-quality males can have higher viability and longer spurs if their quality advantage more than compensates for the loss in viability incurred from having longer spurs. Formally, if a marginally better-quality male survives better by α and makes a greater allocation to spur length which increases his mating success but reduces his viability by β , then mating success and viability will both increase if α is greater than β . Condition-dependent expression of spurs will be favoured as long as the increase in mating success outweighs the net loss in viability. Clearly there is now no real difficulty in explaining why spurs are absent from females — spurs would not increase female mating success but they would reduce female viability.

This natural interpretation may provide the solution to Kirkpatrick's mystery. The equation — longer spurs equals lower

survival — is valid only when male quality is held constant. Once variability in quality is considered, survival and spur length can rise together. This interpretation also adds weight to the conclusion of von Schantz *et al.*¹ that the pheasant data support the 'good genes' theory (sexual preference for high-viability genes). Recent theoretical investigations^{5,6} show that condition-dependent traits are particularly favourable for the evolution of 'good genes' female preference.

But several uncertainties remain about the observations of von Schantz *et al.*. Are male spurs particularly good indicators of differences in male quality? Do female pheasants prefer longer-spurred males in choice experiments when other variables are held constant (for example, territory quality)? And does the variation among males and their offspring have a genetic basis? These questions need to be investigated by experimentation.

ANDREW POMIAKOWSKI

Department of Zoology,
Oxford University,
Oxford OX1 3PS, UK

1. Von Schantz, T. *et al. Nature* **337**, 166–169 (1989).
2. Kirkpatrick, M. *Nature* **337**, 116–117 (1989).
3. Kirkpatrick, M. *A. Rev. Ecol. Syst.* **18**, 43–70 (1987).
4. Nur, N. & Hasson, O. *J. theor. Biol.* **110**, 275–297 (1984).
5. Andersson, M. *Evolution* **40**, 804–816 (1986).
6. Pomiankowski, A.N. *Oxford Surveys in Evolutionary Biology* **7**, 136–184 (1988).
7. Zeh, D.W. & Zeh, J.A. *Am. Nat.* **132**, 454–459 (1988).

Hazards of sulphur

SIR—Meisenhelder and Hunter (*Nature* **335**, 120; 1988) have drawn attention in Scientific Correspondence to the fact that volatile radioactive gas is released from the vial when containers of ³⁵S-methionine or cysteine are first opened. Unfortunately, the situation seems to be even worse.

We have found that ³⁵S- γ -thioATP, which we use for post-translational modification of proteins, also liberates radioactive material and that this material rapidly contaminates any container in which the compound is enclosed. We also have the impression that radioactive volatiles escape from unopened containers as they warm up.

We suggest that the same problem would arise with α -thioATP and perhaps should be considered whenever any radioactive sulphur compound is used. We note that methionine containing a specific pyridine inhibitor is now available and believe that such inhibitors should be widely used by manufacturers. We also suggest that a hazard warning should be printed in all catalogues as well as on data sheets supplied with ³⁵S-compounds.

IVOR SMITH

VANESSA FURST

JOHN HOLTON

Medical Microbiology Department,
University College and Middlesex
Hospital School of Medicine,
London W1P 7PN, UK

Disciplinary institutions

W. F. Bynum

The Investigative Enterprise: Experimental Physiology in Nineteenth-Century Medicine. Edited by William Coleman and Frederic L. Holmes. *University of California Press*:1988. Pp.342. \$39.95.

The American Development of Biology. Edited by Ronald Rainger, Keith R. Benson and Jane Maienschein. *University of Pennsylvania Press*:1988. Pp.380. \$37.95, £30.35.

WHETHER the modern scientific world-view can first be detected in the scientific revolution of the seventeenth century, the secular enlightenment of the eighteenth, or the positivism of the nineteenth, is a matter of historical interpretation and debate. It is more certain that the dominant form of modern scientific communication — the multi-authored paper — is a child of the past century and that it has become unruly of late. In physics, particularly, we are now accustomed to seeing articles with enough authors to stage a football match amongst themselves. And biology is going the same way.

The two fine collections of essays reviewed here are, *inter alia*, about the forces which produced the form of the modern scientific paper. Neither addresses the issue directly, but both are in their separate ways primarily concerned with the practice and the legacy of experimental science in nineteenth-century Germany, whence so many features of contemporary scientific practice derived. A century ago, the scientific paper would probably have been jointly rather than multiply written. But it would generally have described results obtained through experimentation in a laboratory situated in an academic institution. It would increasingly have been published in a specialist journal by authors with allegiances to specialist societies.

The Investigative Enterprise is dedicated to William Coleman, whose recent death was a great loss to historical scholarship. His own chapter — on Jan Evangelista Purkyně and the establishment of the Physiological Institute at Breslau in 1839 — is typical of others in the book in using a particular case-study of a single institution to raise wider issues about the development of what Claude Bernard called experimental medicine. Jacob Henle is the central figure of Arleen Tuchman's analysis of the University of Heidelberg, Carl Ludwig at Leipzig serves a similar function for Timothy Lenoir, and F.L. Holmes uses Justus Liebig, Theodor Bischoff and Carl Voit to look at the programme of metabolic research at the University of Munich. These case studies reinforce our understanding of the extent to which the university in the German-speaking lands was the crucial institution. In France, the academies loomed larger, a

point demonstrated in John Lesch's examination of the Paris Academy of Medicine between 1820 and 1848. Robert Frank's study of the roots within experimental physiology of the electrocardiograph cuts across nations and institutional forms, but underscores a theme common to all of the essays: that clinical medicine and medical education were just as significant for experimental science as the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Founding fathers — Wilson (left) and Whitman of the Marine Biological Laboratory, Woods Hole.

reverse. Indeed it could be argued that, in the nineteenth century, medicine did more for science than science did for medicine. This is not simply because the great experimentalists discussed in *The Investigative Enterprise* came to science through a medical education, but also because the institutions of medicine provided the context within which so much experimental science developed.

That this was so means that the book could almost as well be called *The Entrepreneurial Enterprise*, for nineteenth-century medical scientists had to sell themselves to benefactors, legislators and

hard-nosed educational administrators. Student numbers were just as important then as now, and the fact that German students often spent time at several different universities increased the competitive spirit in the university system and gave scope for innovation in individual disciplines, institutes or research programmes. Examination of these broader professional concerns is splendidly balanced by analysis of the scientific ideas being developed by the key figures. Lenoir's lucid exposition of Ludwig's experimental work on kidney function is marred only by his persistent habit of writing "glomular" instead of "glomerular", and Frank provides an exceptionally clear account of the technical and intellectual aspects of cardiovascular physiology from the sphygmograph to the string galvanometer and the electrocardiograph. In other words, throughout the volume proper attention is paid to what experimentalists actually did with their equipment, especially with those twin pillars of nineteenth-century medical science, the microscope and the kymograph.

In a sense, physiology is merely one part of biology. In practice, by the early twentieth century, the personnel, societies and journals of physiologists were already distinct from those of biologists. From the 1870s, Michael Foster at Trinity College, Cambridge, had taught elementary physiology within the broader context of biology, and his pupil Newell Martin, the first professor of physiology at Johns Hopkins, continued the Foster tradition. For the most part, however, the essays in *The American Development of Biology* explore other facets of biological science and its differentiation into more-or-less autonomous disciplines. Its temporal centre of gravity is a generation later than that of the Coleman-Holmes volume, and its subject matter, restricted to the United States, means that the two books have few individuals in common. Nevertheless, the ideals which powered German experimental science in the mid-nineteenth century still had purchase in what is here argued to be the golden age of American biology. Several key figures, like Charles Otis Whitman and E.B. Wilson, studied in Germany; others, such as Richard Goldschmidt and Jacques Loeb, were Germans by birth. Still others, such as T.H. Morgan and Ross Harrison, were influenced by their study at that most Germanic of American universities, Johns Hopkins.

Without doubt, German achievements inspired pioneer American biologists, but the essays also explore aspects of a story which is rich within its own terms. Sally Kohlstedt and Keith Benson examine the museum origins of biology, while Toby Appel looks at the development of professional biological societies and the unsuccessful attempt of life scientists to

establish a general umbrella organization analogous to the American Chemical Society. Philip Pauly and Jane Maienschein dissect the two principal institutions at which C.O. Whitman worked — the Marine Biological Laboratory at Woods Hole, Massachusetts, and the University of Chicago. These contributions are followed by five case studies of specialization within biology between about 1880 and 1920, a period in which ethology, palaeontology, ecology, genetics and embryology each began to coalesce around key individuals, scientific issues or research strategies.

The collection as a whole is described as the centennial volume of the American Society of Zoologists, which may explain the absence of anything about botany. But it was a happy decision to celebrate a birthday by asking professional historians to write about the wider issues of discip-

line and institution rather than the narrower administrative history of the society itself. Incidentally, the book was published in the year before the centennial. Somebody has been efficient.

Between them, these two volumes represent history of science at its modern best. Scientists will appreciate the extent to which cognitive issues are taken seriously, and historians will value the systematic exploitation of archival material in the placement of the science within its social, disciplinary or educational frameworks. A sociologist might note that, although editorial tasks have been shared, each essay has, with one exception, but a single author: historians of science, it seems, have not yet emulated the people they write about. □

W.F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.

History in science

A.G. Cairns-Smith

Origins: The Darwin College Lectures. Edited by A.C. Fabian. *Cambridge University Press: 1988. Pp. 168. £12.95, \$19.95.*

ETERNAL truths there may be. But to explain the world is not only, and perhaps not ultimately, merely to seek laws of greater and greater generality. Things happen, and some of them leave their imprint. In the choice of connecting theme, these essays express a view of science — both ancient and very much late-twentieth-century modern — in which history has to be taken into account. The unexplained fundamentals in current theories are not only laws of nature, and so on, that just happen to be so, but events that just happen to have happened.

If physics traditionally seeks to concentrate ignorance into a limited set of laws, the next thought is to concentrate ignorance of history into a limited number of critical events or epochs — what we may call origins. Often this has been a way of (negatively) defining a field. Thus traditionally geologists did not have to know about the origin of the Earth, nor biologists about the origin of species or physicists about the origin of anything. Yet that heretic Darwin made sense of biology through speculating about origins (and finding instead a process). Astronomers and physicists are now intent on making sense of 'everything' by speculating about the origin of the Universe — and again finding processes that seem to have brought key features into existence on different time scales. Martin Rees in the opening essay takes us back in imagination to the first 10^{-13} s or so where laws

of physics themselves were coming into being. (Although he warns us that "Theorists differ in how far they are prepared to extrapolate back with a straight face".)

If the origin of our Universe and of our Earth might be said to be singular, with at least most of the creative work having come about in a relatively short time, other origins are much more plural and (like the origin of species) continuing. If I have picked up Ilya Prigogine's message on the origins (plural) of complexity correctly, there are critical but essentially unexplained events which are profusely scattered throughout time and on all scales from the supergalactic to the subatomic. Newtonian eternalism, or huttonian uniformitarianism or darwinian gradualism are not alternatives to the slings and arrows of chaos and catastrophe. They are approximations, useful within limits — places where the view is a little clearer, "windows" as Prigogine might say. Science can no longer do with defining away the rest of reality as its province because it is inconvenient.

Here too are hints as to how the 'hard' and 'soft' sciences will be integrated. They all (even physics) share elements of the eternal, the uniform, the gradual, the chaotic, the catastrophic. In the book we have a remarkably coherent set of essays, based on a series of popular lectures, which bring in physics, chemistry, the Earth's prehistory (David Hughes), and the evolution of man (David Pilbeam), of social behaviour (John Maynard Smith), of society (Ernest Gellner) and of language (John Lyons), as if they were all different parts of the same subject. Well, they are if you take science to mean, simply, trying to understand the world. □

A.G. Cairns-Smith is in the Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK.

Animal rights and the apocalypse

Mary Ann Elston

Animal Liberators: Research and Morality. By Susan Sperling. *University of California Press: 1988. Pp. 247. \$19.95.*

SINCE the mid-1970s, the 'animal rights' movement has developed into one of the most vigorous forms of social protest in Britain, the United States, Australia and many other countries, particularly those of northern Europe. Underlying its diverse manifestations is a common challenge to conventional moral boundaries, a challenge to the assumed rights of human beings to use and, by implication, abuse other animals to further human interests. Why such widespread concern over animals' moral status and suffering in general, and over animal experimentation in particular, has emerged in the late twentieth century in some Western societies is a question social scientists have been slow to take up. Susan Sperling's is one of the first book-length attempts to do so.

Sperling is an anthropologist whose academic career exemplifies to an unusual degree her discipline's concern with both the social and the natural world, and with the boundaries between them. She abandoned physical anthropology for primate behaviour studies, and then cultural anthropology for this study of how the boundaries between the social and the natural world are constructed in her own society and by her own discipline. Her concern is with animals as symbols, mediating human relationships with nature. Protest movements for 'animal rights' are, for her, not fundamentally protests about the suffering of non-human animals. Rather, they are protests about the human condition. Profound anxieties about scientific manipulation of the social and moral order are projected onto animals, specifically vivisected animals as both literal and figurative victims of scientific exploitation. Vivisection thus becomes the supreme evil in an apocalyptic vision of the future and abolishing it the key to achieving a harmonious millennium. The gulf between such zealots and those who seek merely more humane treatment of animals, let alone those who defend current scientific practice, is immense.

This thesis is based on a comparison of the American (or rather the San Francisco) 'animal rights' movement of the early 1980s with the earlier apogee of organized antivivisection protest in late-nineteenth-century Britain. Both periods, argues Sperling, were ones in which prevailing natural cosmologies were disturbed by the expansion of technological capacity and diffusion into popular culture

of new theories about the relationship between humans and other animals: darwinism then, primatology now. She finds other parallels in the social composition of both movements and in the ideological linkages with other contemporary social movements, notably feminism.

The thesis is a bold but essentially simple one. Although sometimes diminished in effectiveness by too much repetition, much of what she writes is illuminating. But her supporting evidence is disappointing. Her account of the victorian movement covers relatively well-worked ground, and is based primarily on secondary sources. She occasionally displays a worryingly cavalier attitude to the fine details, especially of chronology, and greatly exaggerates the direct effect of experimental physiology and bacteriology on nineteenth-century clinical practice and the uniformity of beliefs among antivivisectionists.

The account of the modern movement is based mainly on interviews with nine representatives of the grass-roots groups

campaigning against local universities, and on analysis of their pamphlets. This narrowly drawn sample has led her to concentrate unduly on one element in the modern movement and hence exaggerate its millennial character and its similarities with the earlier movement. In her search for the over-arching symbolic significance of the emergence of animal rights in Western technological societies (or rather in some of them), she has neglected the cultural specificity, complexity and heterogeneity of both the historical and the contemporary movements. Her rigid distinction between the humane movement and 'animal rights' activism, and her identification of modern antivivisection with apocalyptic views of the Earth's future, may have been appropriate for understanding the Californian movement in the early 1980s. But it provides only a partial understanding of, for example, modern British antivivisectionism. □

Mary Ann Elston is in the Department of Community Medicine, University College London, 66-72 Gower Street, London WC1E 6EA, UK.

Lost connections

Margaret A. Boden

The Artificial Intelligence Debate: False Starts, Real Foundations. Edited by Stephen R. Graubard. MIT Press: 1988. Pp. 311. Pbk \$9.95, £8.95.

PEOPLE wanting an accessible entrée into exciting disputes within artificial intelligence should read this book (a reprint of a recent number of *Daedalus*). But the moralists among them will be saddened, and the literal-minded disappointed. For the promise explicit in the title is not honoured — and the sub-title gives the game away. If a debate is an argument in which each side has equal opportunity to make its case, this collection of essays is no debate: one side is enthusiastically favoured.

The "false starts" adumbrated in the sub-title, so most contributors tell us, were the early attempts of AI workers to model various aspects of intelligence by exploiting the logical-formalist properties of digital computers. The "real foundations", we are repeatedly informed, are the attempts — which have burgeoned in recent years — to ground a theory of intelligence in 'connectionist' computer models, in which massively parallel computations are carried out by myriad individual units analogous to neurons. But as the moralist and the literal-minded would surely have us admit, current connectionist models are 'toy' systems, relying on units which (though they share some important properties with nerve cells) are in many computationally signifi-

cant ways very unlike neurons — and which, as yet, are usually simulated on von Neumann machines, as opposed to being implemented in connectionist hardware.

Certainly, connectionist models have produced some interesting, even surprising, performances. We understand why such systems, in general, are able to do some things (such as noise-tolerant pattern recognition) which brains do superbly but which cannot in practice be achieved by more traditional techniques of computer modelling. And we understand why they have some capacity for learning from example, why they can be 'trained' rather than programmed. But we do not know what can, or cannot, be learnt by different connectionist systems. Nor can we always explain *post hoc* how they were able to perform in a given way (learning to reproduce the sounds of spoken words, for example): we do not know which features of the task were the essential constraints, nor which features of the 'successful' model were crucial. Sometimes, we do not even have strong intuitions about these matters, still less rigorous proofs that our intuitions are well-founded. These theoretical difficulties are not highlighted by the enthusiasts, and an innocent reader would hardly realize that they exist.

Mathematical proofs about the computational potential of various types of connectionist system are interesting irrespective of whether such systems have been implemented. The same was true of Turing's proofs (in the 1930s) about computability, and of the theorems about neural computation offered (in the 1940s) by Pitts and McCulloch. But it is worth noting that their seminal paper, which prompted the

first 'connectionist' work, on perceptrons, also prompted pioneering AI research in the formalist tradition (and had previously influenced von Neumann in his logical design of the digital computer). This fact alone should make one doubt whether there is such a deep paradigmatic divide between the "false starts" and the "real foundations" as most contributors suggest.

This point is made in the opening salvo of the confrontation, the chapter by Papert — the connectionists' demon king. Allowing that there is room within AI for insights drawn from connectionist as well as traditionalist theories, Papert nevertheless notes the small-scale nature of today's connectionist models, surmising that "the entire structure of recent connectionist theories [like the perceptrons he criticized so decisively in the 1960s] might be built on quicksand".

The developing relations between neuroscience and AI are optimistically outlined by Schwarz, although Edelman's insistence that selection, not computation, is the key concept in understanding the brain sits uneasily with both branches of AI research. Cowan and Sharp sketch the history of the mathematical ideas underlying diverse forms of connectionist computation, from perceptrons to Boltzmann machines. Dreyfus heaps scorn on the conjectures investigated by 'traditional' AI over the past 30 years, greeting the recent research on parallel distributed processing with a (largely unfounded) *I told you so!* Sokolowski gives a useful conceptual analysis of the term 'artificial', pointing out that, even if natural and artificial intelligence differ significantly, the study of the latter variety would be part of science (as opposed to technology) if it helped us to understand — whether by emulation or contrast — our natural capacities. Poggio describes advances in the computational theory of vision, based on broadly connectionist ideas. Further essays are contributed by noted workers in and commentators on the field: Waltz, Hillis, McCarthy, Turkle, McCorduck, Putnam and Dennett.

As a readable account of some of the achievements and many of the hopes of current connectionism, this collection is both useful and enjoyable. But it should have dealt more seriously with the shortcomings of today's connectionist systems, some of which arguably call for a sequentialist, even a formalist, theory of certain aspects of mental processing. Popper's advice on how to deal with one's intellectual opponents in the most constructive — and the most telling — way was ignored: state their case (or allow them to state it) as well as possible, before mounting your attack. Only then can you be sure that the beast is rightly slain. □

Margaret A. Boden is in the School of Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QN, UK.

Why simians are smart

Carel P. van Schaik and Jan A.R.A.M. van Hooff

Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Edited by Richard Byrne and Andrew Whiten. Clarendon: 1988. Pp. 413. Hbk £50, \$75; pbk £25.

IN tests, some animals solve problems that seem to require creative insight: they show intelligent behaviour. If we accept the idea, suggested by Charles Darwin, that intelligence is an adaptation, we are forced to ask ourselves "What use is intelligence to animals?". The traditional notion is that it enables them to solve ecological problems, such as remembering where to find food and making tools to obtain it. We now know that most animals in the wild do not use tools very often or very intelligently. For instance, in captivity the orang-utan is the Houdini among animals, whereas in the forest the ape's highest achievement is to snap branches to form a night nest or scratch its back with a twig.

It was not until a more cognitive approach began to eclipse behaviourism that another idea gained prominence. As was argued most forcefully by Nicholas Humphrey, animals use intelligent behaviour particularly in their dealings with their companions. This idea is attractive because the social environment, unlike objects, is reactive, thereby favouring the animals' ability to decide flexibly on the best course of action and providing the right conditions for an 'arms race' of mental capacities. Meanwhile, reports appeared suggesting that monkeys and apes actually use these abilities in their everyday social life by forging and breaking alliances, bargaining for social favours, and even trying to deceive others by using social signals out of context. In short, they seem to behave like astute politicians. This has led primatologist Frans de Waal to draw an analogy with Machiavelli's *The Prince*.

In *Machiavellian Intelligence*, over 20 authors have been brought together in order to assess the state of this fledgling field. The book is something of a hybrid between a reader and an edited volume: almost a third of the material has been published before. The editors have included the classic papers, although some modern readers may be confused rather than enlightened by the early ones. Byrne and Whiten also invited experts to review progress, present new data or take stock of the current situation. By adding a few connecting chapters of their own in which they set out the issues clearly, they have



Limited expectations — East London slum life, 1899. The photograph (from the British Library) is reproduced from Atlas of Disease Distributions: Analytical Approaches to Epidemiological Data by Andrew D. Cliff and Peter Haggett. Although most of the maps are of disease patterns in the late 20th century, the atlas spans over a century. The book, to be published on 9 March by Basil Blackwell, costs £75, \$150.

created a highly cohesive book that will serve as an excellent introduction to social intelligence.

Several chapters are devoted to the seemingly trivial task of documenting the depth of the social knowledge of monkeys, apes and children. In practice, ingenious experiments or controlled observations were required to establish conclusively that monkeys recognize that two other individuals are relatives, or that one of them is dominant over the other. Alexander Harcourt underscores the complexity of the rules used in coalitions and thus supports the idea that intelligence evolved because of the need to make calculated decisions based on the instant evaluation of the fast-changing social scene. These contributions prepare the reader for the following chapters in which primates are shown to be 'natural psychologists' who impute intentions and expectancies to others. Possession of this faculty lets them anticipate the moves of others more easily, but it also facilitates the tactical use of deception. Anyone who has watched primates is convinced that they occasionally deceive one another. Critics, however, will point to the anecdotal nature of much of the evidence, and even if they grant that deception occurs will question whether, in order to explain its most complex manifestations, we really need to assume that some apes think that they know what other apes are thinking. If it turns out that we do need to make such an assumption, then the complexity of deception would provide a yardstick for measuring intelligence.

Some chapters deal with competing ideas. The tool-use hypothesis, although inadequate when applied to primates in general, seems ideally suited to explain the rich material culture in the smartest of all primates. Ironically, the chapters concerned with human beings serve to bolster the social-intelligence hypothesis. Peter Smith shows that children reach certain developmental stages of understanding earlier with respect to people than to objects. And Thomas Wynn shows that during the phase of spectacular expansion of the human brain tools remained much the same.

Katharine Milton also takes an ecological perspective, putting emphasis on the energetic costs of maintaining large brains and thus on the factors that limit intelligence. Unfortunately, the comparative studies required to test these competing hypotheses will be difficult to carry out because we do not really know how to measure social complexity or social intelligence. The problem illustrates the point that these are early days, but at least the social-intelligence hypothesis is proving to be heuristically useful for probing the limits of social intelligence in primates. It is good scientific practice ruthlessly to apply Occam's razor, but in the case of social intelligence the simplest explanations may turn out to be those that assume higher mental abilities in the animals. □

Carel P. van Schaik and Jan A.R.A.M. van Hooff are in the Laboratory of Comparative Physiology, University of Utrecht, PO Box 80086, 3508 TB Utrecht, The Netherlands.

Chaos in the international arms race

Siegfried Grossmann and Gottfried Mayer-Kress

Simple nonlinear models, incorporating novel concepts, can shed light on the build-up of armaments between nations.

'DETERMINISTIC chaos' has emerged as a novel and important concept in the natural sciences. Unexpected features in the time evolution of nonlinear, complex systems have been revealed, and study of them has paved the way to understanding the erratic temporal behaviour of macroscopic variables in terms of surprisingly simple dynamical models. These successes in the natural sciences have stimulated attempts to apply the concept of chaos to other fields. This article, which builds on earlier suggestions of Saperstein¹, investigates the possible relevance of chaos in models for international arms races. We present several simple nonlinear models for the arms race, and discuss the anticipated role of chaos in each. Our view is that of the theoretical physicist—we propose models and derive results from them, but the political interpretation of these results remains open for discussion.

There is now a fair amount of literature that surveys the basic ideas of deterministic chaos both from technical^{2,3} and non-technical viewpoints⁴. So here we restrict ourselves to a few topics that generally do not receive enough emphasis. Our goal is to clarify what chaos is not, as well as what it is.

Dynamical system

Consider a dynamical system, that is, a system that can change in time. Let $x_\alpha(t)$ be the system's variables, where the index $\alpha = 1, 2, \dots, d$, at time t . The dimension d of the dynamical system can take any value, but even for small values—say 3—interesting phenomena can occur. For arms-race models, α could represent nations or organizations of allies; for example, $d = 2$ if we consider NATO and the Warsaw Pact, or just the United States and the Soviet Union. If Europe and China are included as independent variables, then $d = 4$. In regional conflicts other nations are involved, but there are still fairly few dominant variables in the global arms race. In physics $x_\alpha(t)$ often describe the coherent macroscopic dynamics of a system which is composed of many microscopic components, so the elements of $x_\alpha(t)$ are often called order parameters, or modes or amplitudes.

How does $x_\alpha(t)$ behave as a function of time? Most familiar are steady states, in which the dynamical variables eventually approach fixed, constant values: that

is, $x_\alpha(t) \rightarrow x_\alpha^*$ for time $t \rightarrow \infty$. In technical jargon, this steady state is called a *fixed point*. The particular steady state x_α^* depends, of course, on certain control parameters a , which here are policies, strategies, economic constraints and so on. We usually first explore how the steady state $x_\alpha^*(a)$ depends on a and then alter the control parameters to adjust the state according to our demands. Steady states, if stable, make the system completely predictable; in strategic modelling, they produce a calculable, and in this sense safe, future.

One of the great surprises to emerge from studies of nonlinear dynamics has been the discovery that stable steady states are the exception rather than the rule. One class of more complicated behaviour involves sustained, self-generated, stable periodic oscillations, that can be not only sinusoidal but deformed (for example, sawtooth-like) variations of $x_\alpha(t)$. Various ecosystems exhibit such periodic variations; one example is a model for large and small fish in the Adriatic Sea, the celebrated Volterra predator-prey system, for which $d = 2$ (see ref. 5). Although this periodic change in an otherwise static environment seems counterintuitive at first, the behaviour of such systems can still be predicted because the regularity of the cycles makes it possible to eliminate unwanted effects of the x_α -oscillations. However, as business cycles and unemployment rates demonstrate, it can be difficult to regulate such periodic cycles.

A closer look reveals that in many nonlinear systems the $x_\alpha(t)$ often behave in an even more complex manner as functions of time t : namely, they can become chaotic. Although there is no generally accepted definition of chaos, there are several properties that characterize chaotic temporal behaviour of $x_\alpha(t)$. One is that long-term behaviour is unpredictable even though short-term behaviour is controlled by apparently simple, although nonlinear, laws. (See ref. 6 for a discussion of predictability in chaotic systems.) This can have various consequences: the illusory belief that successful short-term management allows total control of a system; difficulty in making a diagnosis from the available short-term data; and application of inappropriate control mechanisms that can actually produce the opposite of the desired effect. Is an observed temporal

change of $x_\alpha(t)$ part of a bounded, irregular, chaotic fluctuation from which the system will soon recover, or does it mark the onset of a global instability, the outbreak of a conflict? It takes much more time to measure reliably the mean future development of the system when it is in a chaotic state. This can lead to overshooting or undershooting, and the response becomes either too frantic or 'too little too late'. Inappropriate measures may cause the fluctuations to resonate and become unstable.

Attractive behaviour

But the central point is that although they may be chaotic, the variables in a dynamical system will nevertheless usually be bounded, and the behaviour they exhibit will be 'attractive'. More technically, for chaotic solutions $x_\alpha(t)$ generally stays within finite bounds, and it reapproaches that bounded range if small (short-term) external perturbations have temporarily put it outside. Such a situation, although showing chaotic time behaviour, is 'stable' in the sense that typical solutions will remain bounded and well defined.

There are many examples of such stable chaotic states in physical systems. For instance, the heat conductivity of a fluid layer varies chaotically within a few per cent of the mean conductivity, and the voltage drop at a semiconductor shows some five per cent amplitude chaos. In addition, there is evidence that the economic fluctuations and unemployment rates of the United States are irregular, chaotic signals, but nobody would hesitate to call the system stable in a global sense as long as the fluctuations remain bounded. The recent global dynamics of the stock market is one example.

An interesting feature of nonlinear systems, and one that might be relevant in arms-race models, is transient chaos. Here, the system eventually settles down to a regular behaviour but before then its dynamics might show all the properties of chaos. The most popular examples of such behaviour are games of fortune, such as roulette, where eventually the ball always reaches a simple fixed point. In arms-race models we are not really interested in the long-term behaviour (that is, thousands of timesteps) but more in the dynamics of, say, the next 30 years. A way to recognize this transient chaos is to estimate the effect

of small perturbations on the dynamics: if the perturbations have a large effect on the solutions, that is, if we observe a large spread in the possible histories, then we would assume that we have some sort of transient chaos (see for example ref. 12).

Chaos, therefore, should not be confused with instability. In particular, regarding the arms race and international relations, it is wrong to identify the general onset of bounded chaos with the outbreak of a war or other global crises. The really dangerous case is *instability*. By this we mean that the $x_a(t)$ leave their previously bounded range rapidly and dramatically. The ultimate result may be yet another stable state, itself perhaps chaotic, periodic or even a fixed point. On the other hand, instabilities may well destroy the system, as in the 1929 stock-market crash.

What we have really learned from the theory of chaotic systems is that nonlinear laws simultaneously determine various aspects of a system's temporal behaviour. Not only are there many 'phenotypes' of stable states—fixed points, cycles and chaotic motions—but these may coexist in the same system (for different initial values, for example) and each of them may be stable or unstable. Further, the nature of their stability can change as the control parameters are changed.

A more precise summary of what chaos entails will explain why chaos should not be identified with the outbreak of war. Several features of chaotic motion can be described simply in nontechnical terms: (1) the $x_a(t)$ continue to vary in time, despite static boundary conditions and static environmental control parameters; (2) the range of the x_a -variation is bounded, indicating some degree of global stability; (3) the time-dependence is non-periodic and erratic; (4) although completely determined by a set of differential equations, the $x_a(t)$ cannot be predicted in the long run more precisely than by the above-mentioned bounded range of variability.

Other features

Other equally important features of chaos must be expressed in more technical language. First, the details of chaotic behaviour depend sensitively on the initial conditions or on disturbances. This sensitivity is caused by an internal trajectory instability (positive Lyapunov exponent(s)), a characteristic property of the solutions of the equations of motion under the given conditions. Second, the frequency spectrum $|\tilde{x}_a(\omega)|^2$ is broad, with or without additional sharp peaks. A combination of periodic and superimposed chaotic temporal behaviour may be present. Third, correlations decay, that is, one cannot control the details of the future development from any temporally limited experience or data set.

The presence of chaos, observed as large-amplitude noise generated by the nonlinear couplings of the x_a , makes it harder to analyse the system's development and predict its response to the control mechanisms. But chaos does not necessarily mean loss of stability. On the contrary, standard stability analysis has to be applied even in chaotic states, although this analysis is more complicated mathematically than for fixed points or cycles.

When is chaos likely to occur? There must be at least three dominant continuous variables $x_a(t)$ of the continuous time t . Systems with two such variables cannot be chaotic (for example, refs 2, 3), although they can be globally unstable. This mathematical statement seems to have its correspondence in history, where relations between two superpowers are more stable than relations between many powers of equal importance. If three or more nations are participating in an arms race, additional complexity arises from the possibility of alliances being formed. Each of the participating nations now not only has to react to its opponent's policies but also has to decide with whom to form an alliance. This is a situation which might result from major arms reductions by the two superpowers, as discussed in the strategic arms reduction talks (START).

In the context of a generalization of Richardson's models, this has been discussed in ref. 13. There are indications that there might be political configurations under which the behaviour of the three-nation system shows strong sensitivity to small external perturbations. The corresponding uncertainty about alliances can also be interpreted as a form of instability even if the basic deterministic dynamics is not chaotic. If systems are not continuous in time, they are more liable to become chaotic. For instance, if there is delay in information processing or if the dynamics becomes discrete—perhaps because decisions are made at discrete intervals—then even one-dimensional laws of motion can show chaos.

Small-scale chaos seems to be a feature of the daily affairs of both individuals and nations. (By small-scale chaos we mean irregular fluctuations of the overall dynamics with an amplitude of less than 10 per cent, of the order of 1 per cent, say.) On a microscopic (personal) level, such small-scale fluctuations can be dramatic: losing a job or going bankrupt is always serious for the individual affected, but for the overall economy these events often act as stabilizing fluctuations. Depending on the situation, small-scale fluctuations can help stabilize the system by reducing the probability of catastrophic changes, or the fluctuations can drive the system into a new and qualitatively different state. However, neither result is unique to chaotic states; they are also evident in regular periodic cycles. The

essential property here is the nonlinearity of the equations, not the form or behaviour of the solutions.

The problem is to keep the small-scale fluctuations under global control, or at least to recognize a potential destabilization of a system in which small shocks might induce catastrophic changes. One of these danger signals can be 'critical slowing down', in which it takes an increasingly longer time for the system to recover from small perturbations.

Impending instability

Although the correlations usually decay rapidly, the correlation decay time generally becomes extremely long if the system approaches instability; this critical slowing down is thus an indicator of impending instability. We are only now beginning to understand the stability properties of chaotic states of nonlinear systems (for an abstract example, see ref. 7). In the rest of this article, we shall concentrate on the interplay of chaos and instability in models of the arms race.

Because data on Gross National Product, armaments and so on are available mainly through statistics that are published annually, it is tempting to model the temporal development of security-related quantities using discrete nonlinear dynamics. The first such model was developed by Saperstein¹.

Consider two competing nations. Let x_i and y_i denote the respective fractions of the available resources that nations X and Y devote to armaments. Say t is discrete and labels the successive years. Saperstein's *Ansatz* is

$$\begin{aligned}x_{i+1} &= 4ay_i(1 - y_i) \equiv f_a(y_i) \\ y_{i+1} &= 4bx_i(1 - x_i) \equiv f_b(x_i)\end{aligned}\quad (1)$$

It seems reasonable to assume that a nation's armament x_{i+1} will be determined by the hostile armament y_i of the previous year; less reasonable is the factor $(1 - y_i)$, because it reduces the efforts of X to zero if Y has the largest possible ($y_i = 1$) armament.

Equation (1) is basically one-dimensional with two-year steps, because $x_{i+2} = f_a(f_b(x_i)) \equiv f_{a,b}(x_i)$, and thus x_{i+2} is a function of x_i only, and the same is true for Y with $f_{b,a}(y_i)$. Both nations are simultaneously in a steady state $x^*, y^* = f_b(x^*)$, or they simultaneously bifurcate to a periodic state, or they enter chaos together. Saperstein estimated the values for the control parameters a and b for several pairs of nations, using data on the x 's and y 's for two successive years before the Second World War (compare with ref. 1). His basic thesis was that the appearance of chaos heralded the outbreak of war.

Various aspects of the model are open to criticism: the mathematical assumption itself, the difficulty of acquiring data for

a and b , the choice of the variables x and y and their definition, and the absence of a check on consistency of the model if more than two successive years are considered. The model raises two questions. First, why should the onset of chaos when a and b increase be identified with the outbreak of war; second, does reducing the model to one dimension restrict it too severely?

The answer to the latter question is 'yes'. The conclusions drawn from one-dimensional systems may be far too optimistic (see below). The first question, however, requires fuller consideration. In fact, the problem discussed by Saperstein¹ is that of the loss of stability of a given state, and he makes no use of chaos and its properties. Although Saperstein might have intended differently, his analysis supports our assertion that arms-race models should be studied with respect to stability and not simply with respect to the existence of chaotic solutions. Equation (1) turns out to be stable, but that is because of the one-dimensional oversimplification.

One observation, connected with chaos, is worth noting. When increasing a control parameter forces a system to enter a chaotic state, one of several outcomes typically takes place before chaos sets in (compare with refs 2, 3): (1) period doubling with universal properties; (2) quasi-periodic behaviour; (3) intermittency, that is, irregular but limited bursts of chaotic signals.

Period doubling is present in equation (1). It serves as a warning that chaos may appear when control parameters are further changed. These transition models have been studied extensively but do not themselves represent the fully chaotic solutions.

As we mentioned earlier, the question of stability of a state is nearly always determined by linear stability analysis, even when the state is chaotic and the analysis has to be modified. Stability in arms-race modelling has been thoroughly analysed by Richardson⁸⁻¹⁰. His discussion, although improved by modern developments, has not been invalidated. Our models (presented below) are extensions of Richardson's insights, and are influenced by recent progress in discrete nonlinear dynamics (compare with refs 2, 3).

Our first generalization of Richardson's model is

$$\begin{aligned}x_{t+1} &= x_t - k_{11}(x_t - x_s) + k_{12}(x_m - x_t)y_t \\ y_{t+1} &= y_t - k_{22}(y_t - y_s) + k_{21}(y_m - y_t)x_t\end{aligned}\quad (2)$$

Here x_t and y_t again denote the fraction of the available resources that were devoted to armaments in nations X and Y during the year t . The change of military expenditure by X ($x_{t+1} - x_t$) is taken to be coupled to the military effort y_t of Y. Richardson multiplied y_t by a constant

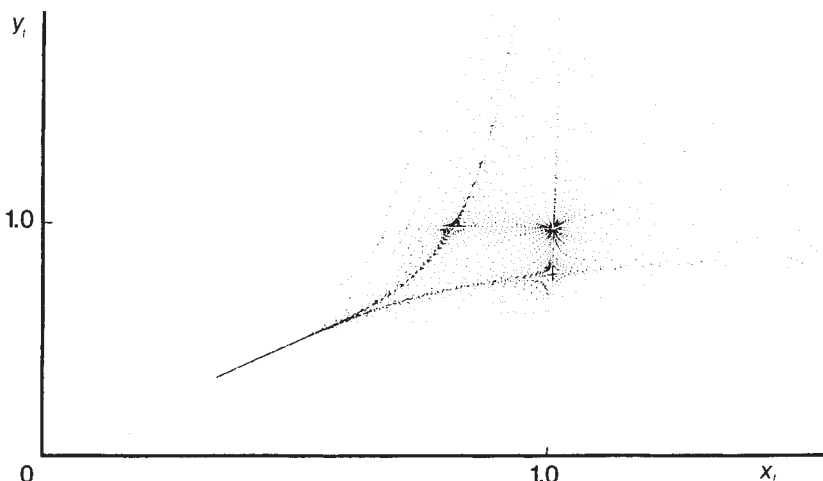


Fig. 1 The x - y plane of equation (3) for $k_{11} = k_{22} = \kappa = 1.0$, $k_{12} = k_{21} = k^{-\frac{1}{2}} = 0.7$, $x_s = y_s = 0.1$, $x_m = y_m = 1$. The figure shows the position of the fixed points [$P_1 = (1, 1)$, $P_2 = (1, s + \frac{k}{\kappa})$, $P_3 = (s + \frac{k}{\kappa}, 1)$, $P_4 = (\frac{\kappa s}{\kappa - k}, \frac{\kappa s}{\kappa - k})$] through the orbits of several different initial conditions. If it starts close to the unstable fixed point P_1 ($x_0 < 1$ and/or $y_0 < 1$), the system first approaches one of the hyperbolic fixed points P_2 or P_3 and is then attracted by the stable fixed point P_4 . Note that the relevant region lies within the unit square, $x \leq 1$, $y \leq 1$. The approach to the stable fixed point is predominantly along the symmetry axis $x = y$ (weakly stable manifold).

factor k_{12} , but we assume that this factor itself depends on the remaining available resources and so we write it as $k_{12}(x_m - x_t)$ for X and $k_{21}(y_m - y_t)$ for Y. The largest available fraction of resources (x_m, y_m) can be set at 1 by re-scaling the equation of motion (2). The positive parameters k_{12} and k_{21} correspond to Richardson's "defence intensity". The variables x_s and y_s are the nations' natural self-establishing armament levels, corresponding to Richardson's "grievance". The positive expense coefficients k_{11} and k_{22} force the approach to x_s and y_s in the absence of hostile nations.

The problem with more refined models is that they require more parameters, for example six in equation (2) compared with two in equation (1). More data are needed to estimate the parameters and so to apply the model. With considerable effort, Richardson gathered much data and obtained order-of-magnitude estimates for the parameters. Current statistical yearbooks help in collecting data. Also, we can improve the relevance and applicability of a model by introducing more parameters; in particular, we can obtain a fit to data over longer periods of time.

Equation (2) can be considered as a discrete version of Richardson's description involving a nonlinear extension of his defence matrix. We know that both nonlinearity and discrete values of t can lead to local instabilities and ultimately to chaos. In the model this is an effect of overshooting, but evidently it is actually present in real arms races, partly because of the discontinuous data flow and mechanisms of political decision-making.

What can be learnt from equation (2)? For reasons of symmetry, we might treat both nations alike and choose $x_s = y_s \approx 0.1$, $k_{11} = k_{22} \equiv \kappa \approx 0.4$, and $k_{12} = k_{21} \equiv k$. The model is still two-dimensional, unless

both nations happen to start their arms race at the same level, that is, unless $x_0 = y_0$. Then $x_t = y_t$ for all times t , and equation (2) can be simplified to $x_{t+1} = f(x_t)$. (In this case, f is a quadratic function of x_t .) For small values of k we have a stable fixed point; increasing the defence intensity k leads to the period-doubling scenario preceding chaos. Chaos sets in at about $k \approx 3$, with the threshold depending, of course, on the self-expense coefficient κ , which determines the build-up rate in the absence of an external threat.

The range of stability for k in the case $x_0 = y_0$ turns out to be unrepresentative. If we allow for different initial military expenditure, that is, if $x_0 \neq y_0$, the fixed-point solutions generated by equation (2) not only become unstable for much lower k (for $k = 1.6$ in this example), but also the route to instability is quite different. Computer iterations show that for $k < 2 - \kappa$ the solutions approach a stable state. As soon as $k > k_c \equiv 2 - \kappa$, the trajectories escape to $-\infty$, that is, there is loss of global stability. The initial point (x_0, y_0) determines which one of many coexisting attractors is approached only at the marginal value k_c . One finds various periodic cycles, but also chaos.

The resource-limited model in equation (2) has three striking features. First, it generally shows a sharp transition to instability, with no preceding period doubling to signal the threatening instability. Second, the threshold k_c to instability is much lower, if both nations are independent and, not unrealistically, are assumed to react in precisely the same way. This lower threshold demonstrates the importance of using the appropriate number of variables (here $d = 2$). Last but not least, the steady-state solutions (x^*, y^*) are hyperbolic if $k > k_c$, that is, the trajectories first seem to approach the steady state but

then leave that region and accelerate to infinity. This behaviour may be interpreted either as the outbreak of war or as a total break-down of the validity of the model.

One might argue that defence efforts are limited not only by the available Gross National Products but also by a smaller maximum acceptable internal level (x_s, y_s). Therefore, we propose the following globally limited model

$$\begin{aligned}x_{t+1} &= x_t + [-k_{11}(x_t - x_s) + k_{12}y_t](x_m - x_t) \\y_{t+1} &= y_t + [-k_{22}(y_t - y_s) + k_{21}x_t](y_m - y_t)\end{aligned}\quad (3)$$

As before, x_s and y_s are the natural armament levels in the absence of a hostile nation. The constants k_{ij} have the same interpretation as before. The variables x_m and y_m again denote the nations' largest available military budget. Once that money is spent, no further change is possible ($x_{t+1} = x_t = x_m$, and $y_{t+1} = y_t = y_m$). Without loss of generality, we can put $x_m = y_m = 1$ by measuring x_t in terms of x_m , measuring y_t in terms of y_m , and properly redefining the k_{ij} .

Steady states

At this point, the reader will ask whether the future of X and Y described by equation (3) can be predicted without using a computer. It is easy to find the steady states. Let us restrict ourselves, for convenience, to two nations of equal strength, and define the common self-armament level as $x_s = y_s =: s$, the self-expense coefficient as $k_{11} = k_{22} =: \kappa$, and the mutual defence coefficient as $k_{12} = k_{21} =: k$. The fixed-point solutions are

$$\begin{aligned}P_1 &= (1, 1) \\P_2 &= \left(1, s + \frac{k}{\kappa}\right) \\P_3 &= \left(s + \frac{k}{\kappa}, 1\right) \\P_4 &= \left(\frac{\kappa s}{\kappa - k}, \frac{\kappa s}{\kappa - k}\right)\end{aligned}$$

Of course, the P_i 's have a meaning only if they belong to the unit square $0 \leq x, y \leq 1$. Hence the particular fixed point P_1 is always meaningful: it represents the state of maximum military effort of both nations. The other possible fixed points, $P_{2,3,4}$, are either all meaningful or none are. The condition for allowed states (within the unit square) reads $k \leq k_c$, $k_c := \kappa(1 - s)$. At k_c all the fixed points coincide. P_4 is on the angle bisector between the self-armament (s, s) and maximum armament ($1, 1$), whereas P_2 and P_3 are on the right or upper edge of the domain.

Can these steady states be attained? This is decided by linear stability analysis. The only degenerate eigenvalue at P_1 is $\lambda^{(1)} = 1 + k_c - k$. Any (negative) deviation ($\delta x_t, \delta y_t$) from P_1 shrinks or grows by the

factor $\lambda^{(1)}$. Therefore, if $k > k_c$, the state of maximum armament is stable unless k exceeds $2 + k_c$. The other states P_i are not only meaningless in this range but also turn out to be unstable.

P_4 is an attracting steady state only if it is meaningful, that is, if $k < k_c$. It has two eigenvalues ($\lambda_{1,2}^{(4)}$) and two attracting eigenvectors, and $P_{2,3}$ always have (for $k < k_c$ as well as for $k > k_c$) one attracting and one repelling direction; thus they are never attained under realistic (fluctuating) conditions (Fig. 1). We have carried out numerical simulations to study the more complicated asymmetric model $k_{12} \neq k_{21}$. These simulations indicate that the fixed point remains stable in the asymmetric case, the coordinates of the fixed point shifting to lower values with increasing asymmetry. Using the symbolic manipulation code MACSYMA, we followed the stability of P_4 as $k := \max\{k_{12}, k_{21}\}$ and κ were varied. The bifurcations from the fixed point P_4 occurred at complex parameter values, and we could find no attractors with oscillating dynamics.

The simple global resource-limited model in equation (3) seems to behave as we would predict from the history of conflicts between nations. If the mutual defence intensity k is small, reflecting confidence in the other nation's peaceful intentions, there is a rather low level of military expenditure (only slightly above the self-caused level s), which increases with k (range $0 < k < k_c$). If k exceeds k_c (range $k_c \leq k < 2 + k_c$) the nations spend the highest possible amount on arms, irrespective of their internal demands. An arms race is unavoidable if the mutual defence factor k is not below k_c . The change from low-level expenditure on armaments to the high-level one occurs continuously with k , which means the change is more or less undetected by the public. If k gets too large (range $k > 2 + k_c$) the arms race becomes unbounded (but recall we have neglected economic constraints): no finite stable fixed point exists. Thus the condition for the danger threshold, beyond which we observe instability, is given by $k_c = 2 + k_c$.

The model predicts exponentially increasing values of x_t and y_t for $k > k_c$. This is not possible by the very definition of the variables x and y , which represent the fraction of the resources, but, of course, the model loses its validity in the unstable region. To study this region, the model would have to be modified.

The most relevant quantity is $k_c = \kappa(1 - s)$. To avoid a critical situation, the mutual defence coefficient k should be less than the self-expense coefficient κ ; the larger the 'natural' self-armament level s , the smaller k has to be. This demand might be very hard to meet. If s approaches 1, there is practically no escape from maximum expenditure on armaments, as k_c gets so small that k easily exceeds it.

If mutual fear prevails, resulting in $k > 2 + k_c$, instability is also inevitable.

Large expenditure

Our model suggests that we should not be too disappointed to find ourselves permanently in a state of large military expenditure. At least the state of largest armament, $x^* = 1 = y^*$, is stable if k is sufficiently below k_c . Using published data¹¹, we estimated the model parameters for the United States. First we let $s \approx 0.1$, making $k_c \approx 0.9\kappa$. By inserting the data for x_t, y_t we obtained $k \approx 0.81\kappa$ during the late 1970s and $k \approx 0.85\kappa$ for 1987. The United States, therefore, is not yet at the largest possible armament level, but the latter value of k is nonetheless fairly high, with $P_4 \approx (\frac{2}{3}, \frac{2}{3})$.

Model equations and scientific notions such as chaos and instability should not make us overlook the important roles of psychology, fanaticism and accident in the international competition in armaments. When these factors intervene, the k_{ij} increase rapidly, or war breaks out.

We conclude by quoting Richardson⁸:

"Humanist: Do you seriously believe that you can prove, by mathematics, anything about the behavior of nations?"

Author: All that can be proved by mathematics is that certain consequences follow from certain abstract hypotheses. But, after such proofs are established, we shall compare the observed behavior of nations with some of the deduced consequences and so form an opinion as to whether the hypotheses are a true description of nations.

Humanist: But how do you know what hypotheses to take as your starting point?

Author: By intuition, luck, or laborious search."

One of us (G.M.-K.) thanks W. Beyer for help with MACSYMA. We also thank W. Beyer, D. Campbell, J. D. Farmer, D. Holm, R. Kupperman, G. Papcun, P. Stein, B. Wendroff and P. White for comments on the manuscript.

1. Saperstein, A. M. *Nature* **309**, 303-305 (1984).
2. Schuster, H. G. *Deterministic Chaos: An Introduction* 2nd edn (VCH, Weinheim, 1987).
3. Bergé, P., Pomeau, Y. & Vidal, C. *Order Within Chaos: Towards a Deterministic Approach to Turbulence* (Wiley, New York, 1986).
4. Gleick, J. *Chaos: Making a New Science* (Viking, New York, 1987).
5. Haken, H. *Synergetics: An Introduction* (Springer, New York, 1977).
6. Farmer, J. D. & Sidorowich, J. *Phys. Rev. Lett.* **59**, 845-848 (1987).
7. Grossmann, S. Z. *Phys.* **B57**, 77 (1984).
8. Richardson, L. F. *Arms and Insecurity* (Boxwood, Pittsburgh, 1960).
9. Richardson, L. F. *Nature* **135**, 830-831 (1935).
10. Richardson, L. F. *Nature* **136**, 1025 (1935).
11. "World Armaments and Disarmament" in *Stockholm International Peace Research Institute Yearbook* 1986, p. 243 (Oxford University Press, 1986).
12. Saperstein, A. & Mayer-Kress, G. *J. Conflict Resolution* **32**, 636-670 (1988).
13. Mayer-Kress, G. *Proc. Ann. Conf. AAAS*, San Francisco, 1989, in the press.

Siegfried Grossmann is at the Fachbereich Physik der Philipps-Universität, Renthof 6, D-3550 Marburg, FRG. Gottfried Mayer-Kress is at the Santa Fe Institute and the Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

Evidence from neodymium isotopes for mantle contributions to Phanerozoic crustal genesis in the Canadian Cordillera

Scott D. Samson*, William C. McClelland*, P. Jonathan Patchett*, George E. Gehrels* & Robert G. Anderson†

* Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

† Geological Survey of Canada, 100 West Pender, Vancouver V6B 1R8, Canada

Sm-Nd and Rb-Sr isotopic studies of several Phanerozoic orogenic belts have shown that much of the crust in these belts is composed of reworked, pre-existing continental crust. In contrast, isotopic data collected from two of the largest terranes in the Canadian Cordilleran orogenic belt, the Alexander and Stikine terranes, show that these tectonic fragments are composed of mantle-derived continental crust. In many respects the style and rate of crustal accretion of these and related terranes appear similar to those of some Proterozoic crustal regions.

ISOTOPIC studies of Phanerozoic orogenic belts such as the Himalayan, Caledonian and Hercynian have shown that the bulk of the crust in these belts is reworked or remobilized pre-existing crustal material¹⁻⁸. This is in contrast to many Proterozoic belts that have been shown to be largely composed of mantle-derived, juvenile crustal material⁹⁻¹². The juvenile crust in these Proterozoic belts represents new continental-crustal material, not just the reprocessing of continental crust previously extracted from the mantle. The contrast between Proterozoic orogenic processes and those of Phanerozoic time has been further shown in a Rb-Sr and Sm-Nd isotopic study of worldwide granitoids¹³. This study showed that a large amount of an older crustal component must have been involved in the genesis of many Phanerozoic granitoids. Most of the Precambrian granitoids, however, have initial ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd ratios similar to mantle values and therefore no significant older crustal component was involved in their genesis.

The above studies reinforce the idea that only a small amount of mantle-derived, juvenile crustal material has been added to the continents during Phanerozoic time. If this is correct it implies that other major Phanerozoic orogenic belts should also be composed of mostly reworked, pre-existing continental crust. Sm-Nd and Rb-Sr isotopic studies of other major orogenic belts, such as the North American Cordillera, would show to what extent these regions are similar to the previously studied Phanerozoic orogens.

Because the North American Cordillera is composed of many tectonostratigraphic terranes the isotopic characterization of this orogenic belt can be accomplished only by the characterization of the individual terranes. The isotopic study of an allochthonous terrane can determine to what extent its formation and accretion represents the true crustal growth of a continent and to what extent this process represents the emplacement of a pre-existing crustal fragment. Here we consider the Nd and Sr isotopic characterization of two of the major terranes of the Canadian Cordillera, the Alexander and Stikine terranes (Fig. 1a).

Alexander and Stikine terranes

The Alexander terrane is one of the largest terranes in the North American Cordillera. It extends over 1,000 km, from Coastal British Columbia, through southeastern Alaska to the Saint Elias Mountains in the Yukon Territory and eastern Alaska (Fig. 1a). The Alexander terrane has been associated with the Wrangell terrane since at least 309 Myr BP¹⁴ and both terranes have been near the Peninsular terrane since at least late Jurassic time^{15,16}. Early-Middle Jurassic deformation along the eastern margin of

Alexander has been interpreted as its initial juxtaposition with North America¹⁷. Final structural accretion of Alexander to the western margin of the Stikine terrane occurred during mid-Cretaceous time¹⁸.

The displacement history of the Alexander terrane is not well constrained. Palaeomagnetic data suggest that the terrane was in equatorial regions for much of the Palaeozoic¹⁹. The presence of bivalves of low palaeolatitudinal affinity in Triassic rocks of the terrane supports this inference²⁰.

Magmatically and tectonically active phases of the Alexander terrane can be grouped into three main periods. The first phase, which lasted from the earliest Cambrian to early Devonian time, was characterized by extensive arc-like volcanism and plutonism and by two orogenic events, the Wales orogeny and the Klakas orogeny²¹. The terrane was tectonically stable through most of the Middle Devonian to mid-Permian time. The second active phase was marked by Triassic volcanics and sediments that are interpreted to have formed in a rift environment^{21,22}. The third active phase occurred in the early Cretaceous with the intrusion of granodioritic plutons which may have been associated with an active volcanic arc.

The Stikine terrane is the largest allochthonous terrane in the Canadian Cordillera¹⁶. It extends from near the United States/Canada border to the Yukon Territory/Alaska border (Fig. 1a). Stikine rocks range in age from probable Devonian metavolcanics to Jurassic subaerial and submarine volcanic and sedimentary rocks²³.

By about Middle Jurassic time the Stikine terrane and the Cache Creek terrane, a terrane along the eastern margin of Stikine, are considered to have accreted to North America^{18,24}. The displacement history of the Stikine terrane, like that of the Alexander terrane, is controversial. Palaeomagnetic data from Permian and Triassic rocks do not show any significant latitudinal transport of the terrane with respect to the stable North American craton^{25,26}. It is possible that the terrane evolved near, but seaward of, the margin of North America.

Isotopic characterization of the terranes

To characterize an allochthonous terrane or any crustal fragment isotopically, representative samples of all major lithological units must be analysed. This includes plutonic, volcanic, metamorphic and fine-grained clastic sedimentary samples. Fine-grained sediments are important because the geological processes of sedimentation can produce good averages of exposed crust and Sm and Nd are not significantly fractionated by the sedimentary process^{27,28}.

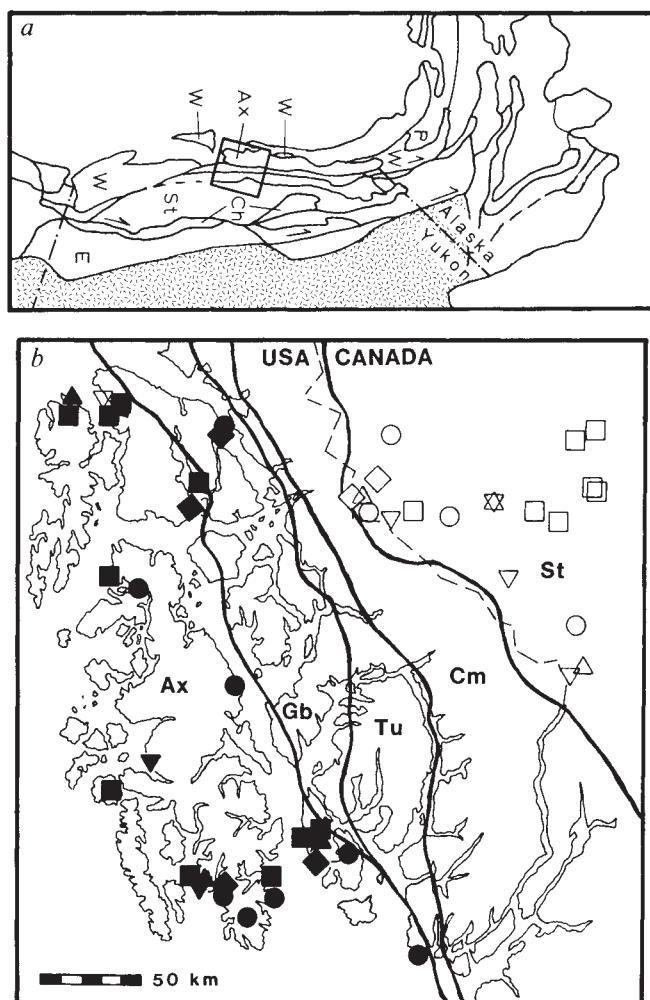


Fig. 1 *a*, Tectonostratigraphic terrane map of the North American Cordillera (adapted from ref. 15). Terranes labelled as follows: Ax, Alexander; St, Stikine; W, Wrangellia; Ch, Cache Creek; P, Peninsular; E, Eastern assemblage of terranes (includes Quenel terrane). Area studied shown by square. *b*, Sample locality map (enlargement of area in *a*). Closed symbols are locations for the Alexander terrane samples, open symbols are locations for the Stikine terrane samples. Triangles, felsic volcanics; inverted triangles, mafic volcanics; circles, felsic plutonics; diamonds, mafic plutonics; squares, sediments. Terrane boundaries are shown by heavy lines. Ax, Alexander terrane; Gb, Gravina belt; Tu, Taku terrane; Cm, Coast Mountains; St, Stikine terrane.

Twenty-eight samples of the Alexander terrane, including Cretaceous plutons intruded into the terrane, were analysed for both Nd and Sr isotopic compositions (Table 1). Although two of the Cretaceous plutons were found in the late Jurassic/early Cretaceous Gravina assemblage (see Fig. 1*b*), they have been included in this study because they probably interacted with the Alexander terrane at depth, and are therefore probes of that terrane. Of the Alexander samples, ten were fine-grained clastic sediments. Plutonic samples of the terrane include mafic rocks such as gabbro; intermediate compositions such as quartz monzonite; and felsic rocks such as granite and trondhjemite. Volcanic samples range in lithology from basaltic to rhyolitic. The sediments analysed are mudstones, siltstones and fine-grained sandstones. The range in age of the samples is from >560 Myr to ~90 Myr. Sample location sites are shown in Fig. 1*b*.

Twenty-two samples of the Stikine terrane were analysed for Nd and Sr isotopes (Table 1). The range in lithologies of the Stikine samples is very similar to that of the Alexander samples. Of the eight sediment samples analysed, six are Middle and

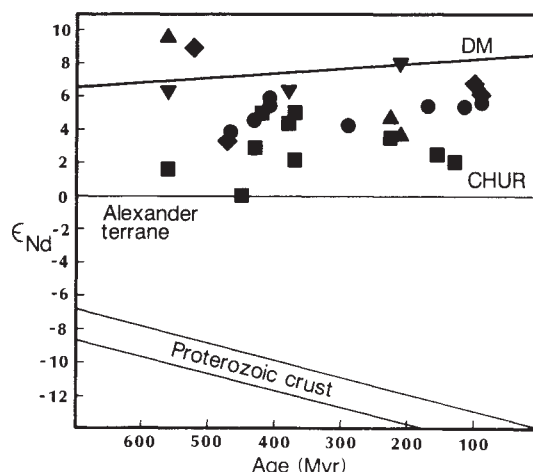


Fig. 2 Plot of initial ϵ_{Nd} against age for the Alexander terrane samples. Symbols as in Fig. 1. Shown for reference are the bulk Earth evolution line (CHUR), depleted-mantle curve of DePaolo⁹, and a Proterozoic crustal evolution line ($^{147}Sm/^{144}Nd = 0.115$, $\epsilon_{Nd}(1.8 \text{ Ga}) = +3.9$). The high positive ϵ_{Nd} values indicate the primitive nature of the Alexander crust.

Upper Jurassic sediments from the Bowser Basin, a clastic sedimentary sequence overlying the Stikine terrane²⁹. A quartz monzodiorite of probable Cretaceous/Tertiary age, also not strictly a sample of the Stikine terrane, was analysed because it is emplaced in the terrane and is probably a good probe of Stikine crust. The range in age for these samples is from 400 Myr to ~65 Myr. Sample location sites are shown in Fig. 1*b*. Precise localities for all samples can be obtained from S.D.S.

In addition to the data presented here, a large amount of initial $^{87}Sr/^{86}Sr$ data has been published for the North American Cordillera (see, for example, ref. 30).

Results

The results are given in Table 1 and Figs 2, 3 and 4. Seven repeat Nd analyses (Table 1) were performed to determine reproducibility. Six of the repeat analyses agreed to within $0.2\epsilon_{Nd}$, and the seventh analysis agreed to within $0.5\epsilon_{Nd}$. Rb-Sr systematics are often observed to be disturbed for felsic and intermediate volcanics³¹ and therefore reliable initial $^{87}Sr/^{86}Sr$ ratios cannot be determined. Rb-Sr isotope data for this type of sample were therefore not collected. Further analytical information is given in Table 1. A complete discussion of analytical techniques at Arizona has been given elsewhere³².

The rocks comprising both the Alexander and Stikine terranes have high positive ϵ_{Nd} values and low $^{87}Sr/^{86}Sr$ ratios, determined for the time of magmatism or sedimentation. The range of initial ϵ_{Nd} and $^{87}Sr/^{86}Sr$ for the Alexander terrane is 0 to +9.6 and 0.70276 to 0.70707, respectively. Corresponding ranges of values for the Stikine terrane are -0.5 to +7.7 and 0.70355 to 0.70547. There is no correlation between age or lithology and ϵ_{Nd} value or $^{87}Sr/^{86}Sr$ for the volcanic and plutonic samples of the Alexander terrane (Fig. 2), whereas the oldest and most mafic samples of the Stikine terrane generally have the highest initial ϵ_{Nd} values and lowest $^{87}Sr/^{86}Sr$ (Fig. 3).

Figures 2 and 3 show initial ϵ_{Nd} versus crystallization or stratigraphic age for the samples. The ages of the sedimentary samples have been determined from biostratigraphic data. Shown for reference are the bulk Earth (or CHUR evolution line), the depleted-mantle (DM) curve of DePaolo⁹, and an evolution curve for average Proterozoic crust that separated from this depleted mantle 1.8 Gyr ago with a $^{147}Sm/^{144}Nd$ ratio of 0.115. If the terranes were composed of mostly reworked mid-Proterozoic crust then most of the samples would be somewhere near this idealized curve, or below if the terranes were

Table 1 Sm-Nd and Rb-Sr data for rocks from the Alexander and Stikine terranes

Sample	Age	Sm* (p.p.m.)	Nd* (p.p.m.)	$\frac{^{147}\text{Sm}}{^{144}\text{Nd}}$	$\frac{^{143}\text{Nd}}{^{144}\text{Nd}}$ § measured	$\frac{^{143}\text{Nd}}{^{144}\text{Nd}}$ initial	ϵ_{Nd}	$\epsilon_{\text{Nd}}(\text{T})$	Rb† (p.p.m.)	Sr* (p.p.m.)	$\frac{^{87}\text{Rb}}{^{86}\text{Sr}}$	$\frac{^{87}\text{Sr}}{^{86}\text{Sr}}$ measured	$\frac{^{87}\text{Sr}}{^{86}\text{Sr}}$ initial
Alexander terrane													
Igneous Rocks													
1 Metarhyolite	560	2.70	8.04	0.2034	0.513148 ± 6	0.512402	+9.9	+9.5	11.0	87.3	0.364	0.706215 ± 13	0.70331
		2.74	8.12	0.2046	0.513164 ± 5	0.512414	+10.3	+9.7	11.4	91.4	0.361	0.706193 ± 11	0.70331
2 Metabasalt	560	8.29	31.1	0.1610	0.512848 ± 6	0.512257	+4.1	+6.7	6.72	338	0.0575	0.705046 ± 11	0.70459
		8.22	30.9	0.1608	0.512826 ± 6	0.512236	+3.7	+6.2	6.66	339	0.0569	0.704942 ± 10	0.70449
3 Qtz Diorite	525	2.77	9.81	0.1706	0.513006 ± 5	0.512419	+7.2	+8.9	1.49	160	0.0269	0.702960 ± 11	0.70276
4 Diorite	470	4.11	18.2	0.1363	0.512620 ± 5	0.512200	-0.4	+3.3	77.7	349	0.644	0.708079 ± 11	0.70377
5 Granodiorite	468	2.89	15.4	0.1135	0.512574 ± 7	0.512226	-1.3	+3.7	9.87	79.8	0.358	0.707955 ± 10	0.70557
6 Qtz Monzonite	438	5.53	25.3	0.1323	0.512688 ± 6	0.512309	+1.0	+4.6	51.1	558	0.265	0.705618 ± 10	0.70396
7 Trondhjemite	409	3.29	17.1	0.1162	0.512720 ± 6	0.512409	+1.6	+5.8	24.3	1074	0.0654	0.704037 ± 10	0.70366
8 Trondhjemite	409	0.657	3.84	0.1036	0.512668 ± 8	0.512391	+0.6	+5.5	27.5	713	0.112	0.704412 ± 13	0.70376
9 Mafic tuff	380	7.08	33.1	0.1292	0.512796 ± 5	0.512475	+3.1	+6.4	18.9	220	0.249	0.705694 ± 9	0.70435
		7.07	33.1	0.1291	0.512788 ± 6	0.512467	+2.9	+6.2	19.5	226	0.249	0.705652 ± 10	0.70431
10 Qtz Syenite	290	5.03	37.4	0.0814	0.512640 ± 5	0.512486	0.0	+4.3	45.6	371	0.355	0.705132 ± 9	0.70367
11 Rhyolite	225	8.17	39.4	0.1253	0.512774 ± 4	0.512590	+2.7	+4.7					
12 Basalt	210	2.91	9.15	0.1925	0.513037 ± 4	0.512772	+7.8	+7.9	24.2	1,226	0.0572	0.706797 ± 14	0.70663
13 Rhyolite	210	9.29	55.6	0.1010	0.512695 ± 7	0.512556	+1.1	+3.7					
14 Granite	171	44.1	193	0.1383	0.512847 ± 4	0.512692	+4.1	+5.4	280	27.5	29.6	0.775921 ± 12	0.70388
15 Granodiorite	115	0.932	4.52	0.1247	0.512859 ± 7	0.512765	+4.3	+5.4	26.7	656	0.118	0.703829 ± 10	0.70364
16 Diorite	100	1.16	3.85	0.1824	0.512974 ± 3	0.512855	+6.6	+6.7	24.1	252	0.276	0.704966 ± 11	0.70457
17 Hornblende	90	4.99	15.7	0.1924	0.512940 ± 7	0.512827	+5.9	+5.9	5.50	292	0.0545	0.703467 ± 8	0.70340
18 Granodiorite	90	3.55	16.4	0.1306	0.512891 ± 8	0.512814	+4.9	+5.7	58.2	939	0.179	0.703962 ± 11	0.70373
Sedimentary Rocks													
19 Metasiltstone	560	3.01	13.8	0.1315	0.512480 ± 5	0.511998	-3.1	+1.6	20.7	277	0.216	0.706593 ± 14	0.70487
20 Siltstone	450	2.87	13.4	0.1300	0.512443 ± 5	0.512060	-3.8	0.0	36.6	165	0.642	0.709776 ± 13	0.70566
21 Sandstone	430	4.06	18.4	0.1333	0.512608 ± 6	0.512233	-0.6	+2.9	40.9	269	0.440	0.708452 ± 13	0.70575
22 Mudstone	420	9.04	52.2	0.1048	0.512639 ± 6	0.512351	0.0	+5.0	66.3	751	0.255	0.705645 ± 17	0.70412
23 Siltstone	380	3.35	13.1	0.1544	0.512754 ± 5	0.512370	+2.3	+4.3					
24 Siltstone	370	2.87	12.4	0.1403	0.512614 ± 4	0.512274	-0.5	+2.2	12.7	166	0.220	0.708108 ± 13	0.70695
25 Siltstone	350	2.99	11.9	0.1517	0.512794 ± 5	0.512446	+3.0	+5.1	27.0	60.6	1.29	0.710712 ± 11	0.70430
26 Mudstone	215	3.15	14.9	0.1276	0.512723 ± 7	0.512543	+1.7	+3.6	21.7	112	0.561	0.708790 ± 11	0.70707
									21.7	111	0.565	0.708768 ± 10	0.70704
27 Mudstone	155	4.27	20.6	0.1253	0.512692 ± 7	0.512565	+1.1	+2.5	40.4	365	0.319	0.705617 ± 13	0.70491
28 Sandstone	140	4.04	19.2	0.1271	0.512683 ± 7	0.512566	+0.9	+2.1	60.2	437	0.399	0.706685 ± 11	0.70589
Stikine terrane													
Igneous Rocks													
29 Felsic tuff	400	4.36	20.8	0.1270	0.512751 ± 11	0.512418	+2.2	+5.8					
		4.22	20.0	0.1277	0.512761 ± 6	0.512427	+2.4	+5.9					
30 Mafic tuff	400	2.90	11.1	0.1571	0.512794 ± 6	0.512382	+3.0	+5.1	87.0	1,056	0.238	0.706667 ± 11	0.70531
21 Basalt	340	4.17	15.3	0.1653	0.512828 ± 6	0.512460	+3.7	+5.1	4.05	131	0.089	0.705310 ± 11	0.70488
32 Basalt	260	3.50	15.0	0.1417	0.512878 ± 6	0.512637	+4.7	+6.5	22.9	462	0.144	0.704883 ± 8	0.70435
33 Gabbro	225	3.45	11.7	0.1787	0.513006 ± 7	0.512743	+7.2	+7.7	6.63	281	0.068	0.704608 ± 10	0.70439
34 Volcaniclastic	220	2.61	9.49	0.1665	0.512939 ± 9	0.512699	+5.9	+6.7	7.54	439	0.050	0.703965 ± 8	0.70381
25 Andesite	210	4.15	19.1	0.1315	0.512668 ± 7	0.512487	+0.6	+2.3					
36 Rhyolite	200	5.98	24.2	0.1497	0.512557 ± 6	0.512361	-1.6	-0.4					
		6.12	24.7	0.1500	0.512550 ± 5	0.512354	-1.7	-0.5					
37 Granodiorite	198	3.39	16.7	0.1231	0.512664 ± 15	0.512505	+0.5	+2.4	76.5	518	0.427	0.705548 ± 10	0.70435
38 Diorite	191	3.87	15.5	0.1512	0.512877 ± 5	0.512688	+4.7	+5.8	38.2	470	0.235	0.704330 ± 10	0.70369
39 Qtz Monzonite	190	1.45	6.76	0.1299	0.512803 ± 8	0.512642	+3.2	+4.8	58.6	634	0.267	0.704711 ± 11	0.70399
40 Qtz Monzodite	170	4.20	21.7	0.1170	0.512804 ± 5	0.512674	+3.2	+5.0	57.6	394	0.423	0.704828 ± 11	0.70381
41 Qtz Monzonite	65	2.91	14.1	0.1246	0.512801 ± 6	0.512748	+3.2	+3.8	55.0	390	0.408	0.704979 ± 11	0.70460
Sedimentary Rocks													
42 Greywacke	270	5.00	20.3	0.1489	0.512874 ± 5	0.512611	+4.6	+6.3	15.7	250	0.182	0.704498 ± 11	0.70380
43 Siltstone	190	3.50	16.2	0.1309	0.512769 ± 6	0.512606	+2.6	+4.2	88.4	513	0.498	0.705809 ± 11	0.70446
44 Shale	170	3.83	17.2	0.1350	0.512635 ± 7	0.512485	-0.1	+1.3	49.9	137	1.05	0.708013 ± 14	0.70547
45 Shale	170	3.84	18.0	0.1291	0.512706 ± 6	0.512562	+1.3	+2.8	46.0	56.5	2.35	0.709235 ± 14	0.70355
46 Greywacke	155	2.30	10.3	0.1345	0.512699 ± 4	0.512563	+1.2	+2.4	39.1	338	0.335	0.705868 ± 9	0.70513
47 Shale	155	1.57	9.69	0.0982	0.512629 ± 5	0.512529	-0.2	+1.8	50.5	46.2	3.16	0.711049 ± 10	0.70408
48 Greywacke	155	2.91	12.9	0.1368	0.512682 ± 6	0.512543	+0.9	+2.0	29.0	131	0.640	0.706549 ± 10	0.70514
49 Shale	155	3.53	13.5	0.1585	0.512687 ± 7	0.512526	+1.0	+1.7	30.6	63.5	1.40	0.708167 ± 11	0.70509
Metamorphic Rocks													
50 Biotite Schist	191	4.63	21.4	0.1307	0.512761 ± 6	0.512598	+2.4	+4.0	59.0	353	0.484	0.707178 ± 11	0.70587

$\epsilon_{\text{Nd}} = [(^{143}\text{Nd}/^{144}\text{Nd})_{\text{sample}} / (^{143}\text{Nd}/^{144}\text{Nd})_{\text{bulk Earth}} - 1] \times 10^4$; that is, it is a comparison of the Nd isotope ratio of a sample relative to the present-day bulk Earth (CHUR) value of 0.512638.

* 2σ errors for Sm, Nd, Sr concentrations and $^{147}\text{Sm}/^{144}\text{Nd}$ are $<0.5\%$.

† 2σ errors for Rb concentrations and $^{87}\text{Rb}/^{86}\text{Sr}$ are $<1.0\%$.

‡ Errors are ± 2 standard errors of the mean.

§ Ratios normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$. $^{143}\text{Nd}/^{144}\text{Nd} = 0.511866 \pm 2$ (2 standard errors of the mean) for 52 runs of the La Jolla standard during period of study.

|| Ratios normalized to $^{86}\text{Sr}/^{84}\text{Sr} = 0.1194$. $^{87}\text{Sr}/^{86}\text{Sr} = 0.710247 \pm 8$ (2 standard errors of the mean) for 21 runs of NBS-987 standard during period of study.

composed of reworked crust older than mid-Proterozoic. If the terranes were a mixture of ancient and juvenile crust then the samples would probably be in a broad range between the Proterozoic curve and the DM curve. However, the figures show that for both terranes virtually every sample has an initial ϵ_{Nd}

between that of CHUR and the DM curve.

The sediment samples from both terranes have positive initial ϵ_{Nd} values, ranging from 0 to +6.3, but in general have slightly lower ϵ_{Nd} values than the igneous samples. This has been noted in other isotopic studies characterizing continental crust^{33,34}.

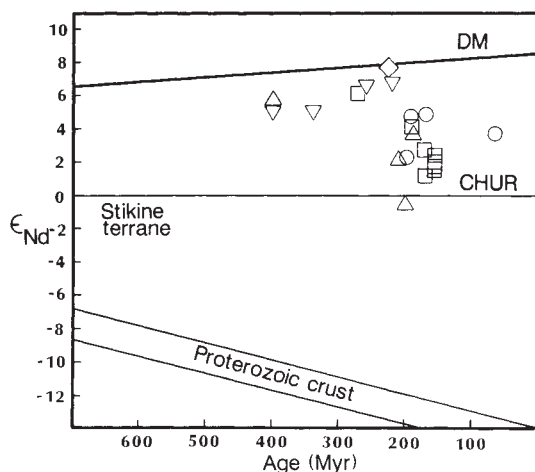


Fig. 3 Initial ϵ_{Nd} versus age for the Stikine terrane samples. Reference lines are the same as in Fig. 2, and symbols are the same as in Fig. 1.

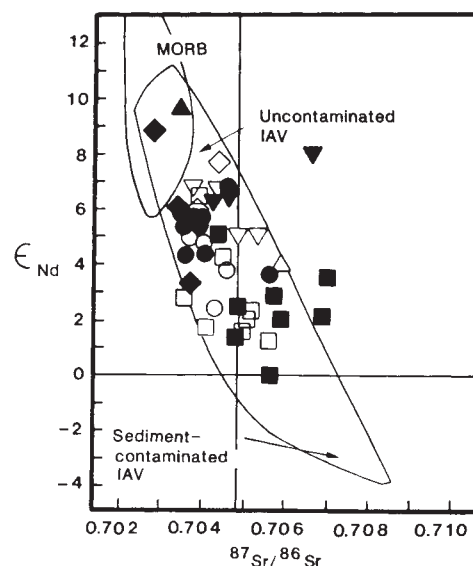


Fig. 4 Initial ϵ_{Nd} versus initial $^{87}\text{Sr}/^{86}\text{Sr}$ for the Alexander and Stikine samples. Shown for reference are the fields for MORB and IAV. Symbols as in Fig. 1. Most of the igneous samples plot near the uncontaminated IAV field.

The lower ϵ_{Nd} value of the sediments can be explained as a mixture of terrane-derived with continent-derived sediment. Just a few per cent of average crustal sediment mixed with terrane-derived sediment could produce the observed values because of the very negative ϵ_{Nd} value of crustal sediment. Another possibility is mixing of a larger amount of sediment with higher ϵ_{Nd} values than those of average crustal sediment, but still lower than pure terrane-derived sediment. Sediment with such intermediate Nd isotopic values, such as $\epsilon_{\text{Nd}} = -4.5$, have been analysed from rivers such as the Columbia, Snake and Merced rivers^{28,35} which drain the western United States. If rivers with suspended material of similar isotopic values existed during the Phanerozoic then sediment from such rivers could have mixed with terrane-derived sediment to give the observed values. Either of these possibilities can also explain the slightly higher $^{87}\text{Sr}/^{86}\text{Sr}$ of the sediment samples (Table 1).

Discussion

The Nd isotopic data show that neither terrane is mainly comprised of older reworked crustal material. Rather, the data clearly demonstrate that unlike crust in some other Phanerozoic orogenic regions, the crust of the Alexander and Stikine terranes is composed almost completely of juvenile, mantle-derived material. Even the sediment samples, although they have lower ϵ_{Nd} values than many other samples, have positive values and show the juvenile nature of the terranes. The juvenile nature of both terranes is further shown in Fig. 4, a plot of initial $^{87}\text{Sr}/^{86}\text{Sr}$ versus initial ϵ_{Nd} . The fields for mid-ocean-ridge basalts (MORBs) and island-arc volcanics (IAVs) are shown for reference. The regions of IAVs uncontaminated by continental sediments (for instance, arcs such as the Aleutians, the Marianas and New Britain) and IAVs heavily contaminated by sediments (parts of the Banda and Sunda arcs, for example) are shown. With the exception of one basaltic sample of the Alexander terrane with an anomalously high $^{87}\text{Sr}/^{86}\text{Sr}$ ratio, which may have been caused by basalt-seawater interaction, most of the igneous samples from both terranes are either in the MORB field or near the uncontaminated IAV field. These data indicate that both terranes probably formed in a volcanic-arc setting, probably in an intra-oceanic environment. The terranes could not have been near a continent during their magmatically active periods—otherwise, some volcanic samples would probably show evidence of continental-sediment contamination, as do modern island arcs such as Banda and Sunda, and would plot near such arc fields (Fig. 4). Similarly, the terranes could not have been obducted upon an evolved continent during any period when they were magmatically active without at least some plutons showing evidence of crustal contamination. Even the

very felsic plutonic samples, such as the Jurassic Bokan Mountain granite body, have very unevaluated isotopic values ($\epsilon_{\text{Nd}} = +5.4$) and display little evidence of crustal contamination.

Most of the sediment samples, although lower in ϵ_{Nd} than the igneous samples, also rule out evolved continental material as a major component of the terranes. The Bowser Basin samples display a general decrease in ϵ_{Nd} value with decreasing age. This is a likely indication of increasing contamination of the terrane sediment by evolved North American sediment.

Phanerozoic crustal growth

Estimates of the proportion of continental crust that has been added to stable continents during the Phanerozoic are quite variable. Some crustal-growth curves suggest that there was almost no addition of crust during the Phanerozoic²⁷. The demonstration that two of the larger terranes of the northern Cordillera are composed predominantly of mantle-derived crust documents new crust added to North America during Phanerozoic time. It is important to determine the rate of addition of this juvenile crust. Determining the addition rates for the terranes is difficult, however, because the duration of their magmatically active periods is not known in enough detail and the amount of material lost to erosion is also unknown. Nonetheless, an estimate of the crustal addition rate can be made by making reasonable assumptions about these parameters. Based on terrane maps of the Cordillera, the area of the Alexander terrane is estimated to be $\sim 100,000 \text{ km}^2$, and the area of Stikine $\sim 300,000 \text{ km}^2$. The crustal thickness of the two terranes is not known. For the purpose of the following discussion the thickness of both terranes, including material lost to erosion, is estimated to be $\sim 30 \text{ km}$. The total volume of the Alexander terrane, therefore, is estimated at $3 \times 10^6 \text{ km}^3$ and the total volume of Stikine is estimated to be $9 \times 10^6 \text{ km}^3$. A conservative estimate of the average ϵ_{Nd} for the terranes is $+4$. Assuming a mantle value of $\epsilon_{\text{Nd}} \approx +7$, $\epsilon_{\text{Nd}} \approx -10$ for reworked Proterozoic crust, and the concentration of Nd in the terranes equal to half that in the reworked crust, the amount of new material in the terranes is slightly greater than 90%. If the period of magmatic activity is taken as 200 Myr for the Alexander terrane, the crustal addition rate is $0.0135 \text{ km}^3 \text{ yr}^{-1}$. If 150 Myr is assumed for the period of magmatic activity of the Stikine terrane, then the addition rate for Stikine is $0.054 \text{ km}^3 \text{ yr}^{-1}$.

Crustal addition rates can also be estimated for the Wrangell,

Peninsular and Cache Creek terranes, terranes which are closely related to Alexander and Stikine. Determining crustal addition rates for these terranes also depends upon the proportion of juvenile crust in the terranes. Although this can only be determined by isotopic studies of the type described here, some evidence exists that most of the crust of the terranes may be juvenile. Most initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for Wrangellia are between 0.703 and 0.705, indicating its probable juvenile nature³⁰. Volcanics of the Cache Creek terrane have ϵ_{Nd} values similar to those of modern arc volcanics³⁶. Assuming, therefore, that these other three terranes are also composed of ~90% mantle-derived material, we can add their crustal addition rates to those of Alexander and Stikine. Estimating the juvenile volumes of Wrangellia, the Peninsular terrane and Cache Creek as $4.5 \times 10^6 \text{ km}^3$, $1.8 \times 10^6 \text{ km}^3$ and $0.72 \times 10^6 \text{ km}^3$ respectively, and assuming that magmatic activity occurred over a period of ~140 Myr, 90 Myr and 90 Myr, respectively, gives an addition rate of $\sim 0.06 \text{ km}^3 \text{ yr}^{-1}$ for these three terranes. The combined rate of crustal addition for all five terranes, keeping in mind the approximate nature of these calculations, is nearly $0.13 \text{ km}^3 \text{ yr}^{-1}$. This growth rate will obviously be larger if other accreted terranes in the North American Cordillera are also determined to be juvenile crustal fragments.

It is interesting to compare the terrane addition rate to more

global addition rates. The Mesozoic–Cenozoic global magmatic-arc addition rate is estimated to be $\sim 1 \text{ km}^3 \text{ yr}^{-1}$ (ref. 37). The terrane rate, therefore, is about 13% of the global arc addition rate. The average continental-crustal addition rate, assuming a present volume of $8.42 \times 10^6 \text{ km}^3$ of continental crust³⁸ generated over 4.5 Gyr, is $1.87 \text{ km}^3 \text{ yr}^{-1}$. This value is higher than the global arc addition rate, but not dramatically different. It is important that the global arc addition rate is similar to the average crustal growth rate and not a small fraction of it, because otherwise the terrane addition rate would not be significant.

It has been suggested that the style of crustal addition in the Canadian Cordillera, growth by terrane accretion, is an important analogue for the style of crustal growth of the late Precambrian Arabian–Nubian shield³⁹. The growth rate of the juvenile material in the central part of the shield has recently been estimated to be ~20% of the global arc addition rate⁴⁰, a value similar to the accretion rate of the Cordilleran terranes. Thus, not only can the pattern of crustal growth of the Cordillera be used as an important model for Precambrian crustal development, but also the rate of terrane accretion can serve as an important analogue for Precambrian crustal evolution.

This work was supported by grants from the NSF and the Petroleum Research Fund. Fieldwork was partially supported by the Geological Society of America.

Received 25 October 1988; accepted 24 January 1989.

- Harmon, R. S. & Halliday, A. N. *Nature* **283**, 21–25 (1980).
- Vidal, Ph. *et al. Phys. Earth Planet. Int.* **35**, 179–190 (1984).
- Ben Othman, D., Fourcade, S. & Allègre, C. J. *Earth planet. Sci. Lett.* **61**, 290–300 (1984).
- Harmon, R. S., Halliday, A. N., Clayburn, J. A. P. & Stephens, W. E. *Phil. Trans. R. Soc.* **A310**, 709–742 (1984).
- Bernard-Griffiths, J., Peucat, J. J., Sheppard, S. & Vidal, Ph. *Earth planet. Sci. Lett.* **74**, 235–250 (1985).
- Frost, C. D. & O'Nions, R. K. *J. Petrol.* **26**, 515–544 (1985).
- Gariepy, C., Allègre, C. J. & Xu, R. H. *Earth planet. Sci. Lett.* **74**, 203–234 (1985).
- Liew, T. C. & Hofmann, A. W. *Contr. Miner. Petrol.* **98**, 129–138 (1988).
- DePaolo, D. J. *Nature* **291**, 193–196 (1981).
- Nelson, B. K. & DePaolo, D. J. *Bull. geol. Soc. Am.* **96**, 746–754 (1985).
- Patchett, P. J. & Arndt, N. T. *Earth planet. Sci. Lett.* **78**, 329–338 (1986).
- McCulloch, M. T. in *Proterozoic Lithospheric Evolution* (ed. Kroner, A.) (Washington, DC, 1987).
- Allègre, C. J. & Ben Othman, D. *Nature* **286**, 335–342 (1980).
- Gardner, M. C. *et al. Geology* **16**, 967–971 (1988).
- Jones, D. L. & Sibirerling, N. J. *US geol. Surv. Open-File Rep.* 79–1200 (1979).
- Coney, P. J., Jones, D. L. & Monger, J. W. H. *Nature* **288**, 329–333 (1980).
- McClelland, W. C. & Gehrels, G. E. *Bull. geol. Soc. Am.* (submitted).
- Monger, J. W. H., Price, R. A. & Tempelman-Kluit, D. J. *Geology* **10**, 7–75 (1982).
- Van der Voo, R., Jones, M., Gromme, C. S., Eberlein, G. D. & Churkin, Jr, M. J. *J. geophys. Res.* **85**, 5281–5296 (1980).
- Tozer, E. T. *Geol. Rundsch.* **71**, 1077–1104 (1982).
- Gehrels, G. E. & Saleeby, J. B. *Tectonics* **6**, 151–173 (1987).
- MacIntyre, D. G. *British Columbia Min. Energy, Mines, Petroleum Resour. Pap.* 1984–1 173–184 (1984).
- Monger, J. W. H. *Can. J. Earth Sci.* **14**, 1832–1859 (1977).
- Cordey, F., Mortimer, N., DeWever, P. & Monger, J. W. H. *Geology* **15**, 1151–1154 (1987).
- May, S. R. & Butler, R. F. *J. geophys. Res.* **91**, 11,519–11,544 (1986).
- Irving, E. & Monger, J. W. H. *Can. J. Earth Sci.* **24**, 1490–1497 (1987).
- McLennan, S. M. & Taylor, S. R. *J. Geol.* **90**, 347–361 (1982).
- Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. *Earth planet. Sci. Lett.* **70**, 221–236 (1984).
- Eisbacher, G. H. *Spec. Publ. Soc. econ. Palaeont. Miner.* **19**, 274–291 (1974).
- Armstrong, R. L. *Spec. Pap. geol. Soc. Am.* **218**, 55–91 (1988).
- Bell, K. & Blenkinsop, J. *Nature* **273**, 532–534 (1978).
- Patchett, P. J. & Ruiz, J. *Contr. Miner. Petrol.* **96**, 523–528 (1987).
- Patchett, P. J. & Bridgwater, D. *Contr. Miner. Petrol.* **87**, 311–318 (1984).
- Patchett, P. J., Gorbatschev, R. & Todt, W. *Precamb. Res.* **35**, 145–160 (1987).
- Goldstein, S. J. & Jacobsen, S. B. *Earth planet. Sci. Lett.* **87**, 249–265 (1988).
- Smith, A. D., Lambert, R. St. J. & Duke, M. J. *M. Geol. Soc. Can. Abstr.* **13**, A114 (1988).
- Reymer, A. & Schubert, G. *Tectonics* **3**, 63–77 (1984).
- McLennan, S. M. & Taylor, S. R. in *The Continental Crust: its Composition and Evolution*, **312** (Blackwell Scientific, Oxford, 1985).
- Kroner, A. *Precamb. Res.* **27**, 277–300 (1985).
- Dixon, T. H. & Golombek, M. P. *Geology* **16**, 991–994 (1988).

The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution

Ravindra Acharya*, Elizabeth Fry*, David Stuart*†, Graham Fox†, David Rowlands† & Fred Brown†

* Laboratory of Molecular Biophysics, Oxford, OX1 3QU, UK

† Department of Virology, Wellcome Biotech, Beckenham, Kent, BR3 3BS, UK

The structure of foot-and-mouth disease virus has been determined at close to atomic resolution by X-ray diffraction without experimental phase information. The virus shows similarities with other picornaviruses but also several unique features. The canyon or pit found in other picornaviruses is absent; this has important implications for cell attachment. The most immunogenic portion of the capsid, which acts as a potent peptide vaccine, forms a disordered protrusion on the virus surface.

FOOT-AND-MOUTH disease remains a major scourge of livestock in several continents. Killed vaccines are available but despite their effectiveness they are difficult to administer in remote regions and severe economic losses due to the disease

still occur. Knowledge of the structure of the virus should assist in the development of improved and novel vaccines and may lead to the design of anti-viral drugs.

Virus crystallography is technically and methodologically testing and the determination of virus structures without the use of heavy atom derivatives is highly desirable. Much work has been

† To whom correspondence should be addressed.

Table 1 Data used in structure determination

Resolution range (Å)	No. unique observed reflections	Percentage of possible observations
10,000–25	43	18.86
25–15	888	92.50
15–10	2,800	95.96
10–7.5	4,758	95.79
7.5–5.0	20,274	96.05
5.0–3.5	51,750	95.39
3.5–3.0	43,101	85.45
3.0–2.9	5,488	39.81
2.9–2.8	1,836	10.73
Total (to 2.9)	129,100	86.98

R-factor (%), 13.9 (all data included). The synchrotron parameters were: Wiggler field strength, 5 T; current, 150–280 mA; energy, 2 GeV; wavelength, 0.85–1.0 Å. Crystal to film distance, 170–190 mm; collimator diameter, 300 µm; oscillation range 0.4–0.5°. Number of crystals examined, 500; number of usable filmpacks (1 filmpack per crystal), 135; number processed, 106.

directed to this end and notable successes include the determination of the structures of Mengo virus¹, where phases were extended from 8 Å using 60-fold non-crystallographic redundancy, and turnip crinkle virus² where 15-fold redundancy and starting phases derived from an homologous virus were used. We report here the successful analysis of foot-and-mouth disease

virus (FMDV) despite a relative paucity of non-crystallographic symmetry (5-fold). The structure revealed provides insight into many properties of the virus.

FMDV belongs to the Picornaviridae, a family of single-stranded positive-sense RNA viruses³. The capsid is composed of 60 copies each of four proteins, VP1–VP4. The proteins VP1–3 are partly exposed on the capsid surface while VP4 is internal and has an N-terminal myristic acid^{4,5}. The picornaviruses fall into four genera^{3,6}: the enteroviruses, for example, poliovirus; the rhinoviruses, for example, human rhino virus (HRV); the cardioviruses, for example, Mengo virus; and the aphthoviruses (FMDV). The rhino and entero genera are the most closely related whereas the cardioviruses occupy the middle ground between these and FMDV⁷. X-ray structures are available for members of the rhino⁸, entero⁹ and cardiovirus¹ genera. Each of the proteins VP1–3 contains an eight-stranded β-barrel very similar in tertiary structure and distribution within the virus particle to the capsid proteins of the small RNA-containing (T=3) plant viruses, suggesting a divergent evolutionary relationship¹⁰. Hypotheses on picornavirus structure and function have been put forward on the basis of the crystallographic studies and these are critically discussed in the light of the structure of FMDV which differs from other picornaviruses in several important respects.

The canyon hypothesis⁸ proposes that the site of cell attachment is at a depression on the surface which is inaccessible to antibody, and that this shields it from immune surveillance. Such depressions have been found on the surface of HRV14,⁸

Table 2 Structural comparisons of the individual proteins of the picornaviruses

Comparison	Protein 1	Protein 2	No. equivalenced residues	Percentage of equivalences		r.m.s. (Å)
				Protein 1	Protein 2	
FMDV	VP1 (FMDV)	VP2 (FMDV)	123 (119)	64 (88)	58 (60)	2.7 (3.4)
	VP2 (FMDV)	VP3 (FMDV)	150 (144)	71 (72)	68 (85)	2.8 (2.9)
	VP1 (FMDV)	VP3 (FMDV)	150 (125)	76 (92)	69 (74)	3.2 (2.9)
HRV14	VP1 (HRV14)	VP2 (HRV14)	148 (144)	54 (68)	58 (59)	3.1 (2.9)
	VP2 (HRV14)	VP3 (HRV14)	156 (158)	62 (64)	66 (87)	2.6 (2.9)
	VP1 (HRV14)	VP3 (HRV14)	181 (153)	66 (73)	77 (84)	3.1 (3.1)
Mengo virus	VP1 (Mengo)	VP2 (Mengo)	144 (143)	56 (64)	60 (60)	3.3 (3.1)
	VP2 (Mengo)	VP3 (Mengo)	161 (156)	67 (65)	70 (86)	2.8 (2.8)
	VP1 (Mengo)	VP3 (Mengo)	178 (160)	69 (72)	77 (88)	3.0 (3.1)
SBMV	A (SBMV)	B (SBMV)	197	99	99	0.6
	B (SBMV)	C (SBMV)	194	97	97	0.9
	A (SBMV)	C (SBMV)	197	99	99	0.7
HRV14 vs Mengo virus	VP1 (HRV14)	VP1 (Mengo)	199 (164)	73 (78)	77 (74)	2.3 (2.8)
	VP2 (HRV14)	VP2 (Mengo)	221 (217)	87 (86)	92 (91)	1.6 (1.6)
	VP3 (HRV14)	VP3 (Mengo)	222 (176)	94 (97)	96 (97)	1.6 (1.8)
FMDV vs HRV14	VP1 (FMDV)	VP1 (HRV14)	169 (131)	88 (96)	62 (62)	2.3 (2.0)
	VP2 (FMDV)	VP2 (HRV14)	201 (194)	95 (97)	79 (79)	1.8 (1.8)
	VP3 (FMDV)	VP3 (HRV14)	213 (164)	97 (97)	90 (90)	1.8 (1.8)
FMDV vs Mengo	VP1 (FMDV)	VP1 (Mengo)	172 (132)	89 (97)	66 (59)	2.4 (2.1)
	VP2 (FMDV)	VP2 (Mengo)	203 (197)	96 (99)	84 (82)	1.6 (1.5)
	VP3 (FMDV)	VP3 (Mengo)	219 (168)	99 (99)	95 (92)	1.6 (1.6)
FMDV vs SBMV	VP1 (FMDV)	A (SBMV)	125 (127)	65 (93)	63 (64)	2.8 (3.2)
	VP2 (FMDV)	B (SBMV)	149 (143)	75 (72)	75 (72)	3.1 (3.2)
	VP3 (FMDV)	C (SBMV)	160 (152)	73 (90)	72 (68)	2.9 (2.7)
FMDV vs HA	VP3 (FMDV)	A (HA)	92 (90)	42 (53)	63 (61)	3.4 (3.4)
FMDV vs HFL	VP3 (FMDV)	HFL (C)	64 (69)	29 (41)	62 (67)	4.0 (4.3)
	VP3 (FMDV)	HFL (V)	69 (67)	31 (40)	67 (65)	3.7 (3.7)

Comparisons performed with the program SHP^{42,43} which uses many of the ideas of Rossmann and Argos⁴⁴, allowing automatic definition of the residues defined as 'equivalent'. Comparisons involving influenza haemagglutinin¹⁸ and immunoglobulin⁴⁵ domains as bench marks are included; the former may be homologous but the latter probably not. Abbreviations: R.m.s. (Å), root-mean-square deviation between equivalenced C atoms. Numbers in parentheses refer to the superposition of the β-barrel domain only. HRV14, rhino virus; SBMV, Southern bean mosaic virus (A, B and C refer to subunits); HA, influenza haemagglutinin (the 'jelly roll' of the globular domain²¹); HFL(C), constant region of the antibody HY5; HFL(V), variable region of HY5⁵¹.

poliovirus⁹ and Mengo virus¹ and there is corroborative evidence for the hypothesis in the case of HRV¹¹. For both HRV14 and poliovirus, the major antigenic sites on the virus are exposed on the surface, as is the major antigenic site of FMDV, the so-called 'FMDV loop' (approximately residues 140–160 of VP1). By contrast with the other picornaviruses, however, this loop alone (as a synthetic peptide) can induce high levels of neutralizing and protective antibodies^{12–14}. Furthermore, part of the loop has been implicated in cell attachment¹⁵. These apparently contradictory properties can be understood in terms of the observed disorder of this structure, but cannot be accommodated within the canyon hypothesis.

Structure determination

We previously described¹⁶ the growth, purification, crystallization and preliminary X-ray diffraction analysis of a virus of serotype 0, strain BFS 1860. The crystals belong to space group *I*23, $a = 345 \text{ \AA}$, with one-twelfth of the virus particle in the crystallographic asymmetric unit (5-fold non-crystallographic redundancy). The small size of the crystals (average dimensions for those used here were $0.12 \times 0.12 \times 0.06 \text{ mm}^3$), coupled with the large number of observable reflections, led to very weak diffraction. In common with many other picornavirus crystals they were also extremely radiation-sensitive, enabling only a single small angle (0.4° – 0.5°) oscillation photograph to be recorded from each crystal, using the SERC Synchrotron Radiation

Source (Daresbury, UK). Data collection was made feasible by the characteristics of the radiation: (i) the extreme parallelism of the beam; (ii) its great intensity; and (iii) the relatively short wavelength (close to $0.9 \text{ \AA}$), all of which contributed to an increase in the yield of data from each crystal. Many crystals were examined using the so-called American method¹⁷ resulting in 135 usable film packs of which 106 were analysed (Table 1). A combination of locally written and CCP4 programs was used to produce an almost complete data-set (full details of the method of structure determination will be presented elsewhere). Despite the weakness of the diffraction, the data are of comparable quality to those obtained for other viruses^{1,8,9} (Table 1). Throughout the structure determination and refinement all measurements were used, as even the weakest reflections provide indirectly important restraints on the phases of the stronger data; to this end negative intensities were truncated by the method of French and Wilson¹⁸.

Phase determination proceeded in several steps: (i) Determination of the virion orientation in the unit cell. In the *I*23 cell the virus must be centred at $(0, 0, 0)$ ¹⁶, and there are thus only two possible orientations of the icosahedral symmetry axes. Calculation of a self-rotation function completely resolved this ambiguity. (ii) Determination of low-resolution phases. A 'composite' virus, 'Rhengo', was constructed by summing the electron density (calculated from the atomic coordinates) for HRV14 and Mengo virus (with appropriate density estimates for the

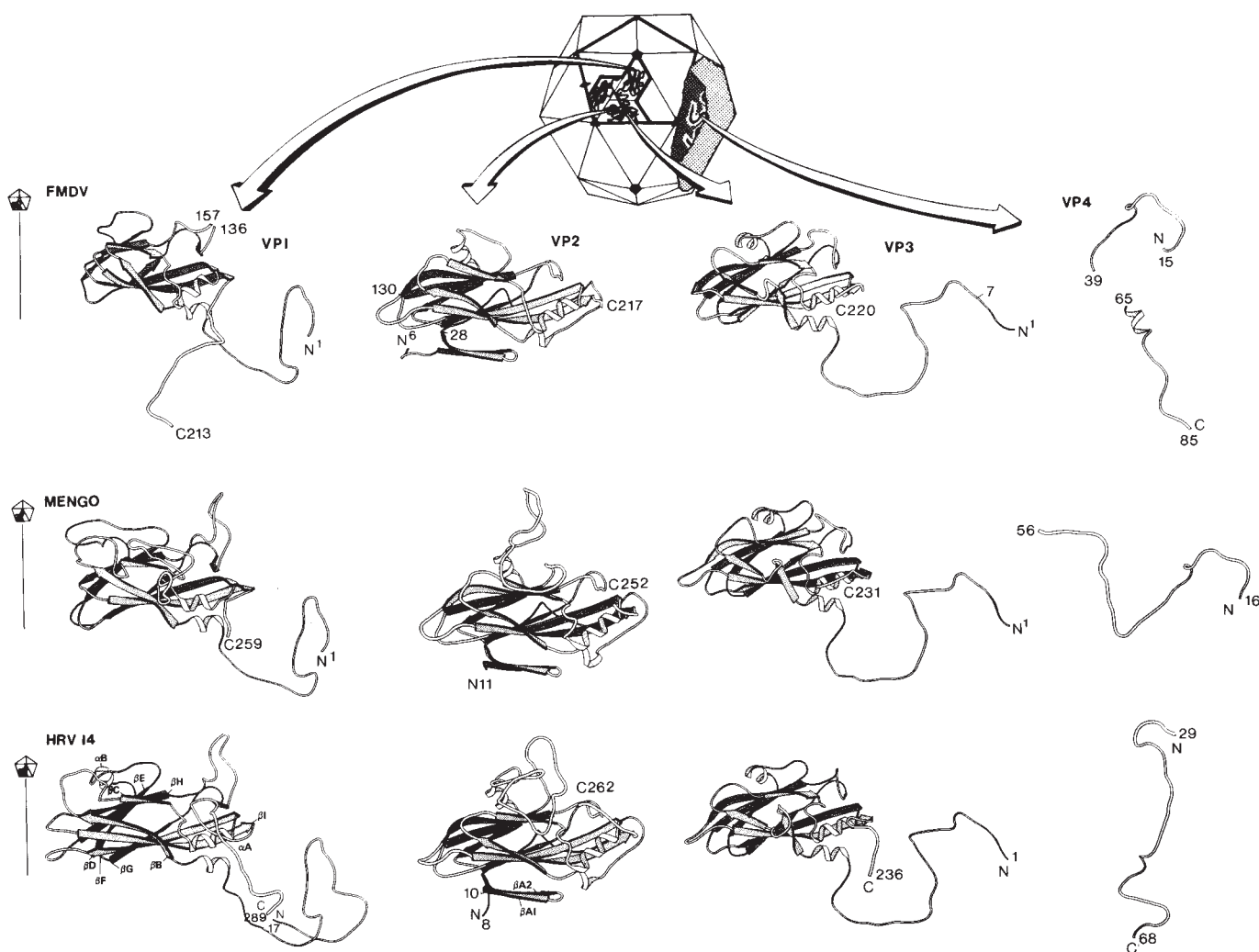


Fig. 2 Picornavirus structure illustrating the disposition of the structural proteins VP1–4 with respect to the icosahedral symmetry axes. The figure shows the detailed folding of each of the proteins in the three picornaviruses FMDV, Mengo virus and HRV14. The appropriate direction of the icosahedral 5-fold axes is shown for VP1. The representation follows Richardson⁴⁶ using a program written by Priestle⁴⁷.

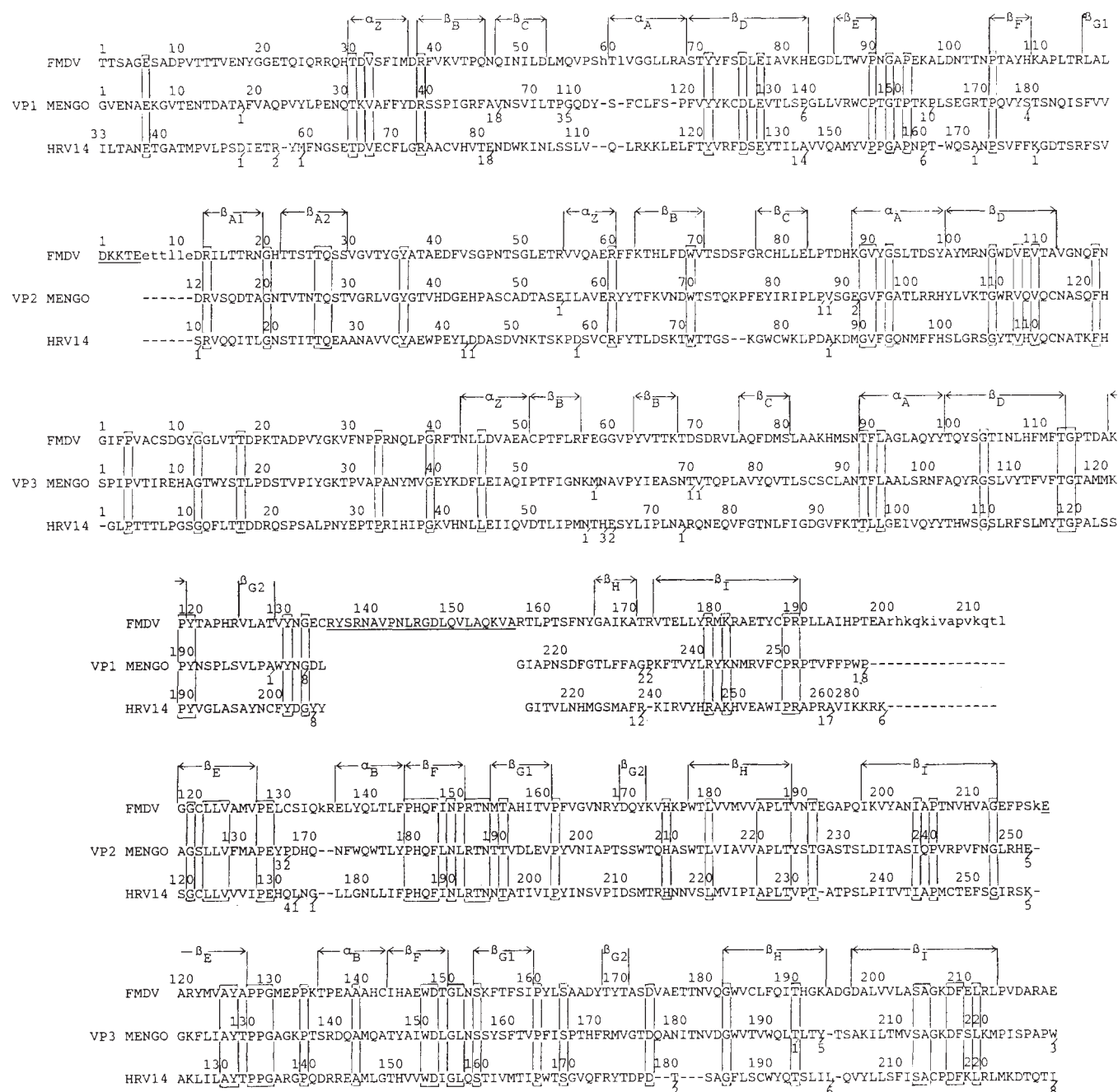


Fig. 3 Alignment of HRV14 and Mengo virus sequences for VP1, VP2 and VP3 to FMDV. Only sequences that have structural equivalence are aligned. A dash represents the lack of a structurally equivalent residue. If this occurs in both HRV14 and Mengo virus then the residue in FMDV is shown in lower-case. Residues in HRV14 or Mengo virus that do not match FMDV are not shown and a number below the line indicates how many residues have been omitted at such points. Sequences underlined are disordered in the X-ray structures and have therefore been omitted from the alignment. Sequences that are identical among HRV14, Mengo virus and FMDV are boxed. The basic secondary structure for FMDV is indicated.

solvent and nucleic acid regions). Fourier transformation of this image gave structure factors that scaled to the observed FMDV data with an *R*-factor

$$\frac{\sum_h |(F_{h,obs} - |F_{h,calc}|)|}{\sum_h |F_{h,obs}|}$$

of 50% for all the data in the range ∞ –8 Å. Deliberate mis-orientation of the virion by 90° increased *R* to 54%. We adopted this simple equal weight strategy because, although sequence comparisons suggested a closer similarity with Mengo virus than with HRV14 (11.4% of the residues of VP1 of HRV14 match

in structure and sequence compared with 14.3% for Mengo virus), the coordinates available¹⁹ for HRV14 were more accurate. (iii) Phase refinement and extension. Iterative molecular averaging and solvent flattening²⁰ followed the methods of Bricogne²¹ and was guided by the experience of Luo *et al.*¹ and Rayment²² in particular. Initial attempts at phase extension failed to produce meaningful phases beyond 8 Å, the average error being 75°. Success was achieved with a novel procedure which included the automatic definition of a tight envelope around the protein shell²⁰ and allowed a much greater volume of the map to be assigned to solvent and RNA (densities for both of which were separately flattened to their mean values).

ar and Cache Creek terranes, terranes which are closely related to Alexander and Stikine. Determining crustal addition rates for these terranes also depends upon the proportion of crust in the terranes. Although this can only be determined by isotopic studies of the type described here, some evidence exists that most of the crust of the terranes may be of juvenile origin. Most initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for Wrangellia are between 1.0705, indicating its probable juvenile nature³⁰. Volcanics of the Cache Creek terrane have ϵ_{Nd} values similar to modern arc volcanics³⁶. Assuming, therefore, that these terranes are also composed of ~90% mantle-derived material, we can add their crustal addition rates to those of Alexander and Stikine. Estimating the juvenile volumes of the Alexander, the Peninsular terrane and Cache Creek as $4.5 \times$

global addition rates. The Mesozoic–Cenozoic global arc addition rate is estimated to be $\sim 1 \text{ km}^3 \text{ yr}^{-1}$ (reference terrane rate, therefore, is about 13% of the global arc rate). The average continental-crustal addition rate, at a present volume of $8.42 \times 10^6 \text{ km}^3$ of continental crust³⁸ over 4.5 Gyr, is $1.87 \text{ km}^3 \text{ yr}^{-1}$. This value is higher than the arc addition rate, but not dramatically different. It is thus that the global arc addition rate is similar to the average growth rate and not a small fraction of it, because the terrane addition rate would not be significant.

It has been suggested that the style of crustal addition in the Canadian Cordillera, growth by terrane accretion, is a good analogue for the style of crustal growth of the Cambrian Arabian–Nubian shield³⁹. The growth rate

Fig. 1 Portions of the electron-density map showing the two regions containing disordered disulphide bridges. The electron density is shown as a three-dimensional net of contours drawn at an arbitrary level. *a*, Density surrounding the hole at the icosahedral 5-fold axis of symmetry showing the S–S bonds between Cys 7 residues of VP3. CA-73 is the α -C atom of Cys 7, and SG73 and SG73A are the sulphur atoms of adjacent Cys 7 residues around the 5-fold axis. *b*, Two discrete conformations of the S–S bond (S and S') between Cys 134 of VP1 (CA-1341) and Cys 130 of VP2 (CA-1302).

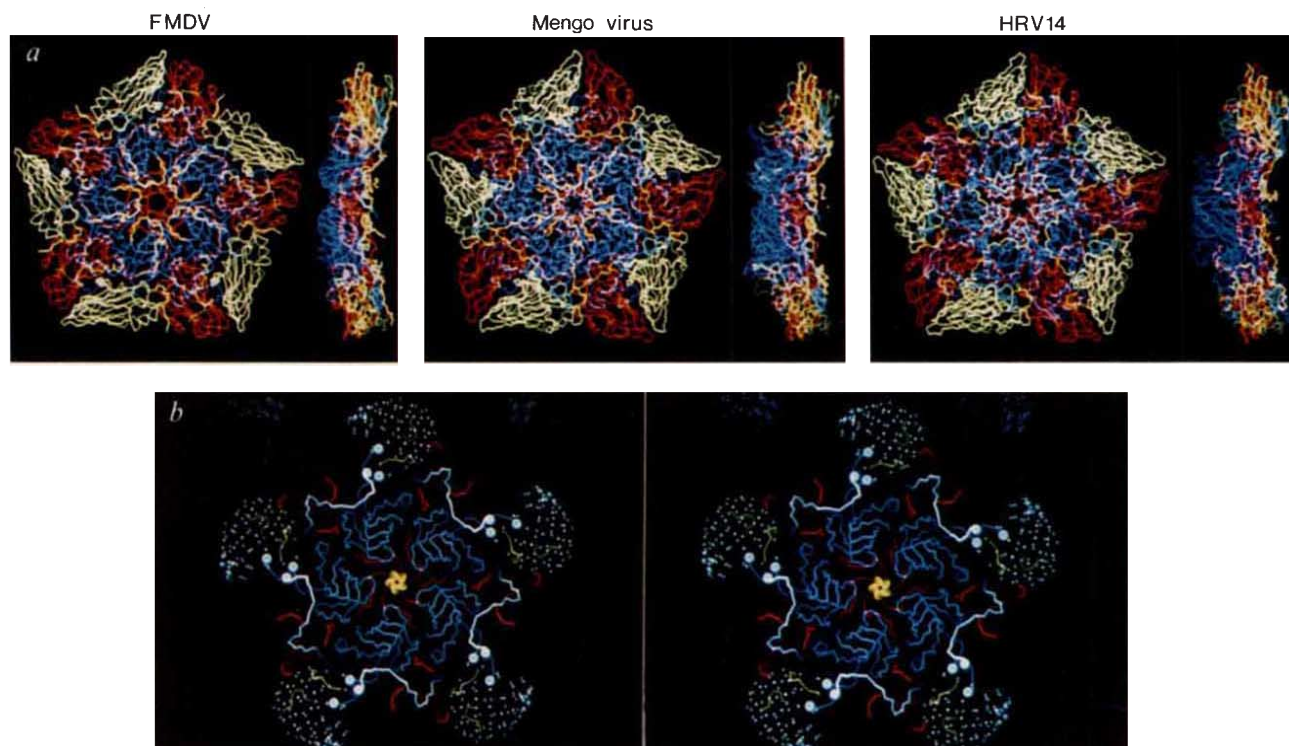


Fig. 4 Pentameric association of promoters. *a*, Orthogonal views of the virtual bonds joining α -carbon atoms for the pentamers of FMDV, Mengo virus and HRV14 respectively. The view of the 5-fold axis is from the outside. Colour coding: blue, VP1; green, VP2; red, VP3; yellow, VP4. *b*, A stereo pair showing the view down the 5-fold axis from the outside. The colour code is as above but additional features are included: position of the disulphide ring around the 5-fold axis, yellow balls; C terminus of VP1 (residue 213) and the ends of the well-defined density for the FMDV loop (residues 136 and 157), blue balls; approximate position of the diffuse density for the FMDV loop, blue dots. The C-terminal portion of VP1 becomes brighter as it passes from its parent across VP3 to the adjacent VP1 subunit.

This resulted in a substantially increased phasing power. Phase extension proceeded very slowly, requiring 519 cycles on a dedicated microvaxII computer and produced maps of extremely high quality with very good statistics. The final averaging R -value was 14.9% on all data from infinity to $2.8 \text{ \AA}$ resolution (compared with 27% for poliovirus⁹) and the correlation coefficient

$$\frac{\sum_h ((F_{\text{obs}} - |F_{h,\text{obs}}|)(F_{\text{calc}} - |F_{h,\text{calc}}|))}{[\sum_h ((F_{\text{obs}} - |F_{h,\text{obs}}|)^2 - (F_{\text{calc}} - |F_{h,\text{calc}}|)^2)]^{1/2}}$$

was 0.92 (compared with 0.84 for poliovirus⁹). The accuracy of the phasing at higher resolution is indicated by the R -factor

and correlation coefficient at $3 \text{ \AA}$, 24% and 0.79 respectively. As the quality of the electron density map is central to our discussion of the structure and its implications, portions are shown in Fig. 1. Note that the phases are derived solely by phase extension with no reference to our model of FMDV.

The sequence assignment and model building were begun at a resolution of $3.5 \text{ \AA}$ (using the program FRODO²³) and were completed with the final $2.9 \text{ \AA}$ map. Over 90% of the residues of the capsid proteins were clearly visible in the electron-density map. The model has been refined using all the data from $5 \text{ \AA}$ to $2.9 \text{ \AA}$ with the program XPLOR²⁴ (R.A., E.F., D.S. and G. Taylor, unpublished observations). The exceptional quality of

the model may be judged from the initial *R*-value which was 32% on all the data and 24% on the stronger data. Refinement reduced these values to 17% and 12% respectively without manual intervention and without the inclusion in the model of bound water molecules. The accuracy of the model may be estimated from the agreement of the five crystallographically independent protomers when refined without the restraint of icosahedral symmetry; the main chain atoms deviated from the mean positions by an average of 0.17 Å.

Architecture of the virus particle

The close structural relationship of FMDV to HRV14 (ref. 8) and Mengo virus¹ is shown in Figs 2, 3 and 4. The overall shape of the well-ordered electron density is that of a spherical shell starting at a radius of 100 Å from the particle centre and extending to an outer radius of 150 Å. There is also a region of diffuse density at a higher radius than 150 Å (Fig. 4*b*). If we ignore this feature (which is discussed fully below) then the outer surface of the virus is relatively smooth. Unlike the other picornaviruses, FMDV does not seem to possess any major canyons or pits; indeed as VP1–3 are substantially smaller in FMDV, the protein shell is generally thinner. As expected from the success of the molecular-replacement phasing, the 'cores' of VP1–3 are placed at essentially the same radius and in a similar orientation to those of the other picornaviruses. VP1 shows the most significant rearrangement. For VP2 and VP3 the axis of the β -barrel is approximately in the plane of the capsid whereas in VP1 it is canted up from the plane towards the 5-fold axis (Fig. 4*a*). This is true of all picornaviruses but is more pronounced in FMDV where there is a further displacement of 1.5 Å. In common with the other picornaviruses, we can see no ordered RNA^{1,8,9}.

Tertiary structure of the capsid proteins

VP1. This protein comprises 213 residues whereas in HRV14, poliovirus and Mengo virus it comprises 279, 306 and 277 residues, respectively). This radical truncation does not result in the loss of any elements of secondary structure (although α B is irregular and not drawn as an α -helix in Fig. 2). The structure is 'trimmed' largely at the loops of the 5-fold axis end of the subunit. This, together with the canting of the barrel, produces a 5-fold axis naked of VP1 loops, in marked contrast to other picornaviruses where the VP1 loops approach close enough to form a putative calcium-binding site⁸. Furthermore, the outer surface of VP1 is sheared of the major excursions that make up most of the walls of the canyon in HRV14. By contrast, the 'inside' surface of VP1 is similar to that of the other picornaviruses.

Although residues 1–132 and 159–209 of VP1 are well-ordered, there are varying degrees of disorder in the rest of the protein. The disordered regions coincide with the major antigenic site (residues 141–160) and part of a minor antigenic site comprising residues 200–213. The region of diffuse density which extends beyond 150 Å radius is associated with residues 137–156 (careful analysis of the very low level electron density has enabled us to place residues 133 to 136, and 157 and 158 at either end of the loop and residues 210–213 at the C terminus).

The 17 C-terminal residues of VP1 form a long 'arm' quite unique among the picornaviruses. This arm is on the surface of the capsid and runs at a constant radius to the 5-fold axis from the VP1 of one sub-unit over VP3 until it nestles among the surface loops of the VP1 of the 5-fold related (clockwise) protomer (Figs 2, 4*b*).

VP2. In FMDV, VP2 comprises 218 residues (whereas in HRV14, poliovirus and Mengo virus it comprises 262, 272 and 256 residues, respectively); of these, the first five cannot be located (this is also the case for the seven and eleven N-terminal residues in HRV14 and Mengo virus, respectively). The electron density for residues 130–132 is rather weak; this is associated

with the disorder of the FMDV loop as residues 134 of VP1 and 130 of VP2 are bridged by a disulphide linkage (see below). As with VP1, loops decorating the outside of the FMDV particle are trimmed compared with those of the other picornaviruses (Fig. 2). The very close association of the N-terminal residues around the 3-fold axis (see below) generates a structure that is not observed in the other picornaviruses.

VP3. In FMDV, VP3 comprises 220 residues whereas in HRV14, poliovirus and Mengo virus it comprises 238, 236 and 231 residues, respectively; it is the most highly conserved of the coat proteins of the picornaviruses (Fig. 2). All 220 residues can be accurately located in the electron-density map.

VP4. This protein varies considerably among the members of the Picornaviridae family. The longest VP4, in FMDV, has 85 residues whereas in HRV14, poliovirus and Mengo virus it has 69, 69 and 70 residues, respectively. Only a little over half of these are clearly defined in the electron density map (residues 15–39 and 65–85) and there is no indication of the position of the myristic acid. The N-terminal portions of VP4 in Mengo virus and FMDV are similar but different from that in HRV14 and lie close to the 5-fold axis. Beyond this region the chain in FMDV forms a well-defined helix of three turns, a novel structure for a picornavirus VP4, before abruptly fading out. There is density for the C-terminal portion of the chain close to the 3-fold axis (Fig. 2).

A hole at the icosahedral 5-fold axis

By contrast with the other picornaviruses, the truncation of VP1 exposes the five-stranded β annulus formed by the N termini of VP3 packing around the icosahedral 5-fold axes. This forms a tube with an average inner diameter of 11 Å. Although VP4 lies below this tube, no well-ordered residues obtrude close to the symmetry axis (in poliovirus the N-terminal residues and myristic acid form an additional inner layer of structure). Thus, a strikingly hydrophobic hole is formed with three highly conserved residues of VP3, Phe 3, Val 5 and Cys 7, forming a constriction which limits the internal diameter to about 6 Å at the narrowest point (Cys 7). The disulphide bridges which link the cysteine residues probably restrict flexibility at this position (see below). The dimensions of the hole are compatible with the penetration of intercalating dyes resulting in photo inactivation of infectivity²⁵ by producing lesions in the RNA²⁶; the RNA within poliovirus is refractory to such inactivation. The hole also explains the rapid penetration of caesium ions, which gives FMDV the highest buoyant density among the picornaviruses.

A putative calcium-binding site

Residues 6, 7 and 8 of VP2 form a tight annulus around the icosahedral 3-fold axes. Within this annulus there is strong density directly on the symmetry axes which is compatible with the presence of a Ca^{2+} ion. The liganding residue appears to be Glu 6 which is conserved in all the picornaviruses. It is conceivable that ionic interactions at this site stabilize the icosahedral association of the pentamers. Although calcium is required for the cell attachment of FMDV²⁷, it is unclear whether the Ca^{2+} ion involved is liganded to the receptor or to the virus (see cell attachment section below).

Acid lability

FMDV is distinguished from other picornaviruses by its instability below pH 7 (ref. 28). In addition to the 3-fold interactions described above adjacent pentamers are held together by several weak interactions notably the series of hydrogen bonds that knit the non β -barrel strands β -A1 and β -A2 of VP2 onto strand β -F of the β -barrel of VP3. These interactions are common to those picornaviruses for which the structure is known. There is, however, a high density of histidine residues lining the pentamer interfaces in FMDV. As the pK value of this amino acid correlates with the pH at which disassembly of the virus occurs, this may provide a qualitative explanation for the observed lability.

Proposed autocatalytic cleavage of VP0

It has been proposed that structurally homologous serine residues in HRV14 and Mengo virus (located at the start of β -A1) are active in the autocatalysis of VP0 to give VP4 and VP2, the final stage in the maturation of the virion^{8,29}. Although there is no structurally homologous serine in FMDV, Arnold *et al.*²⁹ have suggested that one nearby performs the same function. Indeed, there are two serine residues in FMDV (serines 28 and 29 of VP2), which lie close by at the C terminus of β -A2 (Fig. 2). The autocatalysis hypothesis was prompted by the observation in HRV14 of a hydrogen bond between the C terminus of VP4 and the serine. However, within the FMDV particle, these residues approach no closer than 8 Å. Moreover, the first residue of VP2 visible in the electron-density map is at a considerable distance from both serine residues and the C terminus of VP4. Thus, the FMDV structure provides no evidence for the proposal, and there is a strong case for testing the hypothesis by site-directed mutagenesis.

Disulphide bonds

Two-dimensional polyacrylamide gel electrophoresis (PAGE) analysis of the virus proteins shows that a considerable proportion are in the form of dimers linked by disulphide bonds (D.R.). With virus of serotype 0, strain BFS 1860, two forms of dimer are found, one a homodimer composed of VP3, the other a heterodimer of VP1 and VP2. Furthermore, PAGE experiments implicate Cys 134 of VP1 in the heterodimer bridge. By contrast, virus of serotype A, sub-type 10, which does not possess a cysteine residue in that position, has only VP3 homodimers. The electron-density map allows both disulphide bonds to be unambiguously identified. The residue involved in the VP3 homodimers is Cys 7 which lines the hole in the β -annulus at the 5-fold axis. This is a novel structure (Figs 1a and 4b) which has five cysteine residues statistically disordered between 10 possible positions arranged symmetrically around the 5-fold axis. Clearly, at most four out of the five residues can form disulphide bridges around any one 5-fold axis. Although the conservation of Cys 7 of VP3 in all FMDVs that have been sequenced indicates that it has an important function, the virus remains infectious under reducing conditions.

The disulphide bond linking VP1 and VP2 is very close to the major antigenic loop on VP1 and may therefore influence the structure of the antigenic site. The presence of this structural link may explain the analyses with monoclonal antibodies and peptides which indicated that the major antigenic site of serotype 0 viruses contains complex epitopes^{30,31} whereas the antigenic site of serotype A virus has the characteristics of a linear determinant³². Sequences are available for viruses of serotypes A, O, C and SAT3. The combination of cysteine residues at positions 130 on VP2 and 134 on VP1 has been found only in serotype 0.

The major antigenic site

Residues 141–160 of VP1 have been identified as the major neutralizing site on the virus and peptides incorporating this sequence are able to induce high levels of neutralizing antibody^{13,14}. Sequences from the C-terminal region of VP1 can also induce virus-neutralizing antibodies, but to a much lower level¹³. Mapping studies with neutralizing monoclonal antibodies have shown that for virus of serotype 0, residues from both these regions contribute to a single complex epitope^{30,31}. The structure supports this by showing the two regions to be in close proximity on the surface of the virus (Fig. 4b). An unexpected feature is that the C-terminal portion of the epitope in any one protomer is derived from the adjacent protomer, that is, the polypeptide chain wraps across from one protomer to the next (Fig. 4b). Whereas most of the C-terminal region of VP1 is well-ordered, the electron density for the last four residues is more diffuse, indicating disorder in this region.

The main feature of this epitope is the prominent antigenic loop comprising residues 133–158 of VP1. The disorder in this

region of the electron-density map of FMDV means that it has not been possible to assign structure to the antigenic loop although there is a region of low-level 'diffuse density'. From both these data and our previous work, in which VP1 was shown to be intact in crystallized virus¹⁶, it is clear that the loop has not been removed or 'nicked' by proteolysis. The extreme susceptibility to proteolysis indicates considerable flexibility; however, modelling and NMR studies suggest that part of the loop is helical (ref. 32 and P. Mascagni, personal communication).

The interpretation of disorder in a crystallographic experiment is a thorny problem³³. Crystallography yields a time and space average of the structure and from a single experiment we cannot therefore distinguish dynamic from static disorder. Furthermore, errors in the measurement of the structure-factor amplitudes and phases produce a noisy image within which the more diffuse features rapidly become lost. The novel aspects of the phase determination require us to show that we are not over-interpreting our results. We calculated therefore an unaveraged electron-density map with $(2F_o - F_{\text{averaged}})$ amplitudes and α_{averaged} phases. This map is extremely clear and shows some diffuse density for the disordered FMDV loop. At the disulphide position (residue 134 of VP1) two conformations for the sulphur atom can be seen clearly (Fig. 1b); beyond this, the density fades rapidly within the protein envelope. As a further check we have examined an electron-density map with $(F_o - F_c)$ amplitudes and α_{calc} phases, where F_c is derived from the refined model. This map, which cannot be affected by our envelope definition procedure, shows similar disordered density. Despite the difficulty in interpreting the disorder, we can conclude that: (i) the marked protrusion of the loop presents an unusually linear epitope³⁴; (ii) this dominant epitope is loosely connected with the rest of the structure; and (iii) the epitope probably has considerable internal disorder as none of the five independent copies in the virus particle gives a clear image. These features place FMDV in a unique position within the substantial database of known three-dimensional structures and provide insight to why a short peptide might mimic the virus so successfully.

Cell attachment site

Three lines of evidence suggest that the cell-attachment site on the virus involves the FMDV loop (133–158 of VP1) and the C-terminal region of VP1: (i) all antibodies which are known to bind to the loop region prevent cell attachment; (ii) proteolytic cleavage of this loop or the C-terminal region of VP1 abolishes cell attachment; and (iii) short peptides including the highly conserved sequence Arg-Gly-Asp at residues 145–147 are able to inhibit virus attachment to susceptible cells¹⁵. Furthermore, the Arg-Gly-Asp tripeptide has been identified as having cell adhesion properties in other systems³⁵ and BHK cells, which are susceptible to FMDV infection, possess receptors which bind Arg-Gly-Asp. Attachment of FMDV is dependent on Ca^{2+} ions, and known Arg-Gly-Asp receptors have a Ca^{2+} requirement. Indeed the receptors seem to possess classic calcium-binding sequences³⁵ which are, however, deficient in acidic ligands. The Asp residue of the Arg-Gly-Asp sequence could contribute to the Ca^{2+} -receptor interaction³⁵. Such a role might require some flexibility in the FMDV loop. We conclude that the attachment site on FMDV involves at least some of the loop and is exposed on the surface of the virus.

The exposure of residues involved in cell attachment, which are necessarily conserved, to immune surveillance raises the important problem of the observed antigenic variation. As proposed by Colman *et al.* for influenza virus³⁶, the most probable explanation for this is that there is enough amino-acid variation between serotypes around these residues to prevent cross-neutralization by heterotypic antibodies. This hypothesis is supported by the observation that monoclonal antibodies that have been mapped as binding to the Arg-Gly-Asp region of the virus were not inhibited from binding to the virus by peptides with

heterologous sequences flanking the conserved sequence, whereas the homologous peptide and even the tripeptide itself inhibited binding¹⁵. The flexibility of the loop indicates that there are minimal structural constraints on the amino-acid sequence and may be a device to allow for the extensive variation.

The 'canyon' hypothesis is often understood solely in terms of exclusion, that is the dimensions of the putative receptor binding sites in HRV14 and Mengo virus (both deep depressions) prevent accessibility of the base of the site to antibody but allow recognition and binding by a smaller cellular receptor. We propose a more general hypothesis in which 'concealment' of the constant receptor binding site is achieved by 'camouflaging' a small constant region within a sea of considerable variability.

Implications for picornavirus evolution

A comparison of the structures of VP1-3 of FMDV with other β -barrel-containing proteins (Table 2) places Mengo virus roughly mid-way between HRV14 and FMDV in terms of relatedness. Furthermore, the observation that the similarity between the corresponding proteins in different Picornaviridae genera is stronger than between proteins of the same virus¹ is also valid for FMDV. Argos *et al.*³⁷ proposed that the conserva-

tion of cell-binding properties in viruses could account for the conservation of tertiary structure. Although HRV and Mengo virus may attach at homologous sites, recent evidence implicates a different binding site for poliovirus^{38,39} and our observations with FMDV suggest that attachment is at a non-homologous site. The proposal is therefore no longer tenable and the conservation of tertiary structure must reflect the preservation of other properties. These may be structurally dictated by the need to form a stable capsid and to maintain a balance between flexibility and rigidity. This would allow for suitable control of the dynamics of virus assembly and uncoating⁴⁰.

A model for the structure of FMDV based on sequence alignments and the known structures of Mengo virus and HRV14 has recently been published⁴¹. Whereas the model correctly predicts the basic structural features common to all picornaviruses it does not contain the distinctive features of FMDV described above and the sequence alignment on which it was based is in part incorrect.

We thank: Howard Rees and Keith Meldrum; John Mann and David Goodridge; John Helliwell and the staff at Daresbury; Peter Radlett, Barry Penfold and Alistair Telling; Yvonne Jones, Katherine Pelly and Nigel Walker; Garry Taylor; David Phillips and Michael Rossmann; and Lorraine Joyce. The Medical Research Council supported R.A. and E.F.

Received 2 November; accepted 19 January 1989.

- Luo, M. *et al. Science* **235**, 182-191 (1987).
- Hogle, J. M. *et al. J. molec. Biol.* **191**, 625-638 (1986).
- Cooper, P. D. *et al. Intervirology* **10**, 165-180 (1978).
- Chow, M. *et al. Nature* **327**, 482-486 (1987).
- Paul, A. V., Schultz, A., Pincus, S. E., Oroszlan, S. & Wimmer, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7827-7831 (1987).
- Newman, J. F. E., Rowlands, D. J. & Brown, F. J. *gen. Virol.* **18**, 171-180 (1973).
- Palmenberg, A. *Proc. Int. Conf. Virol.* (Newport Beach, California, in the press).
- Rossmann, M. G. *et al. Nature* **317**, 145-153 (1985).
- Hogle, J. M., Chow, M. & Filman, D. J. *Science* **229**, 1358-1365 (1985).
- Rossmann, M. G. *et al. in The Molecular Biology of Positive Strand RNA Viruses* (eds Rowlands, D. J., Mayo, M. A. & Mahy, B. W. J.) 301-320 (Academic, London, 1987).
- Colonna, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5449-5453 (1988).
- Strohmaier, K., Franze, R. & Adam, K. H. *J. gen. Virol.* **59**, 295-306 (1982).
- Bittle, J. L. *et al. Nature* **298**, 30-33 (1982).
- Pfaff, E., Mussgay, M., Bohm, H. O., Schulz, G. E. & Schaller, H. *EMBO J.* **1**, 869-874 (1982).
- Fox, G. *et al. J. gen. Virol.* (in the press).
- Fox, G. *et al. J. mol. Biol.* **196**, 591-597 (1987).
- Rossmann, M. G. & Erickson, J. W. *J. appl. Cryst.* **16**, 629-636 (1983).
- French, S. & Wilson, K. *Acta crystallogr.* **A34**, 517-525 (1978).
- Burnstein, F. C. *et al. J. molec. Biol.* **122**, 535-542 (1977).
- Wang, B. C. *Meth. Enzym.* **115**, 90-112 (1985).
- Bricogne, G. *Acta crystallogr.* **A39**, 823-847 (1976).
- Rayment, I. *Acta crystallogr.* **A39**, 102-116 (1983).
- Jones, A. *Meth. Enzym.* **115**, 157-171 (1985).
- Brunger, A. T. in *Crystallographic Computing 4: Techniques and New Technologies* (eds Isaccs, N. W. & Taylor, W. R.) (Clarendon, Oxford, 1988).
- Galloway, I. A. in *Fifth Progr. Rep. Foot-and-Mouth Disease Res. Committee* 342-345 (HMSO, 1937).
- Brown, F. & Stewart, D. L. *J. gen. Microbiol.* **23**, 369-379 (1960).
- Brown, F., Cartwright, B. & Stewart, D. L. *Biochim. biophys. Acta* **55**, 768-774 (1962).
- Talbot, P., Rowlands, D. J., Burroughs, J. N., Sangar, D. V. & Brown, F. *J. gen. Virol.* **19**, 369-380 (1973).
- Arnold, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 21-25 (1987).
- Parry, N. R. *et al. in Vaccines 85* (eds Lerner, R. A., Chanock, R. M. & Brown, F.) 211-216 (Cold Spring Harbor Laboratory, 1985).
- DiMarchi, R. *et al. Science* **232**, 639-641 (1986).
- Rowlands, D. J. *et al. Nature* **306**, 694-697 (1983).
- Stuart, D. I. & Phillips, D. C. *Meth. Enzym.* **115**, 117-142 (1985).
- Barlow, D. J., Edwards, M. S. & Thornton, J. M. *Nature* **322**, 747-748 (1986).
- Ruoslahti, E. *A. Rev. Biochem.* **57**, 375-413 (1988).
- Colman, P. M., Varghese, J. N. & Laver, W. G. *Nature* **303**, 41-44 (1983).
- Argos, P., Tsukihara, T. & Rossmann, M. G. *J. molec. Evol.* **15**, 169-179 (1980).
- Murray, M. G. *et al. Science* **241**, 213-215 (1988).
- Hogle, J. W. & Filman, D. J. *Proc. R. Soc. (in the press)*.
- Badger, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3304-3308 (1988).
- Luo, M., Rossmann, M. G. & Palmenberg, A. C. *Virology*, **166**, 503-514 (1988).
- Stuart, D. I., thesis, Univ. Bristol (1979).
- Stuart, D. I. *et al. J. molec. Biol.* **134**, 109-142 (1979).
- Rossmann, M. G. & Argos, P. *J. molec. Biol.* **134**, 109-142 (1979).
- Sheriff, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8075-8079 (1987).
- Richardson, J. S. *Adv. Protein Chem.* **34**, 167-339 (1981).
- Priestle, J. P. *et al. EMBO J.* **7**, 339-343 (1988).

LETTERS TO NATURE

A model for soft γ -ray burst repeaters

M. Boer*†, J. M. Hameury‡ & J. P. Lasota§

* Centre d'Etudes Spatiales des Rayonnements, BP 4346, 31029 Toulouse Cedex, France

† Max-Planck-Institut für Extraterrestrische Physik, D-8046 Garching bei München, FRG

‡ Département d'Astrophysique Relativiste et de Cosmologie, Observatoire de Paris, Section de Meudon, 92195 Meudon Principal Cedex, France

§ Département d'Astrophysique Extragalactique et de Cosmologie, Observatoire de Paris, Section de Meudon, 92195 Meudon Principal Cedex, France

Soft γ -ray repeaters (SGRs)¹⁻² represent a new class of astrophysical γ -ray bursts, characterized by rapid but non-periodic recurrence (timescales from 1 s to 1 yr), soft spectra (effective temperature ~ 40 keV), and short duration (~ 0.1 s). They probably cannot be explained by whatever physical mechanism lies behind the more common hard γ -ray bursts: in particular, the thermonuclear model cannot work³. Only three SGRs are known¹,

although some other bursts with soft spectra and short duration might be added to the list, and their spatial distribution and hence distance are undetermined. We regard it as coincidental that two of the three SGRs are close to the plane of our Galaxy. Livio and Taam have argued⁴ that accretion of comets onto neutron stars is consistent with all the observational evidence. We suggest here instead that SGRs bursts are due to comets falling onto a magnetized white dwarf. Our model produces correct spectra and timescales, and has the advantage of solving statistical difficulties inherent in the neutron-star model.

If SGRs are due to the accretion of comets onto neutron stars, one is faced with the statistical problem of why there are so many more such sources than X-ray binaries. SGRs should be rather close⁴ (a few hundred parsecs) and they have long periods of inactivity⁵, so that their density in space would exceed 10^4 kpc⁻³, orders of magnitude larger than that of X-ray binaries (of which there are about 100 in the whole galaxy), which are either their progenitors or their descendants.

In our Galaxy, however, there is a large number ($\sim 20,000$) of magnetic cataclysmic variables (MCVs), whose characteristic hard X-ray spectra are quite similar to those of SGRs. These are semi-detached binary systems, in which a magnetized white

dwarf accretes matter from a low-mass companion. Guided by this similarity, we propose that the SGRs are due to episodic accretion of comets onto magnetic white dwarfs (most probably isolated or possibly descendants of MCVs). The number density of magnetic white dwarfs in the solar neighbourhood is in the range 5×10^{-4} to 10^{-3} pc^{-3} (five magnetized white dwarfs have been observed within 13 pc for declination $\delta \geq -20^\circ$). In addition, unlike the case for neutron stars⁸, magnetized white dwarfs may retain a cometary cloud⁹, as the formation of a white dwarf is a much less violent event than a supernova explosion. The total mass lost during the planetary-nebula phase can be less than 50% of the total mass, allowing for the retention of circular orbits (note that even if the star were to lose as much as 80% of its mass, half of the comets on high-eccentricity orbits would remain bound^{9,10}). Thermal ablation of the comet cloud is also negligible, as it is in the supernova case: even in the extreme limit of all of the flux received by a comet during the whole life of the star ($< 10^{50}/4\pi D^2 \text{ erg s}^{-1} \text{ cm}^{-2}$) being available for evaporating it, the comet would still survive at distances from the star as low as $D = 1,000 \text{ AU}$ (ref. 5). Cooling cannot be neglected, and the effective temperature reaches the critical value of 160 K required to sublimate ice at distance of $\sim 500 \text{ AU}$ for a star luminosity as large as $10^4 L_\odot$; at typical distances of a few times 10^4 AU , the effective temperature for such a luminosity is only 20 K.

Tremaine and Zytow⁵ extensively discuss the proposal of Harwit and Salpeter¹¹ that γ -ray bursts are due to impacts of comets onto a neutron star. We shall adapt their scenario to the case of a white dwarf. The two main differences are that a strongly magnetized white dwarf has a magnetic moment of 10^{33} – 10^{34} G cm^3 , as opposed to 10^{30} – 10^{31} G cm^3 for a strongly magnetized neutron star, and that the gravitational potential is lower at the surface of a white dwarf than of a neutron star. The latter imposes some constraints, discussed below, which imply that the bulk of the emission occurs in the hard X-ray range. The strength of the magnetic moment increases the capture radius of a white dwarf, relative to a neutron star, by an order of magnitude. The comet is disrupted by tidal forces at a distance of $\sim 5 \times 10^9 \text{ cm}$, is ionized by the electric field induced by its motion in the magnetic field at a distance of $\sim 10^9 \text{ cm}$, and then experiences drag resulting from the emission of Alfvén waves. This sequence of events is not sufficient to cause the comet to fall directly on the surface of the white dwarf, but the initially parabolic orbit is transformed into an elliptic one. Accretion occurs on the second (or subsequent) passage⁵; this accounts for the bunching of bursts from SGRs⁴.

The returning fragment is likely to spread over a fraction of the white-dwarf surface, because its ballistic trajectory does not intersect this surface; it is dragged by the emission of Alfvén waves, and destroyed by a Rayleigh–Taylor instability^{12,13}. Such a process is required if the bulk of the energy is to be emitted as hard X-rays, because comets infalling on to non-magnetic white dwarfs are not subject to this instability and penetrate below the photosphere, where their energy is thermalized. From the duration Δt of the γ -ray burst, one can determine the size $l = \Delta t v_{\text{ff}}$ of the fragment producing the burst, where v_{ff} is the velocity of freefall:

$$l = 5.2 \times 10^7 \left(\frac{\Delta t}{0.1 \text{ s}} \right) \left(\frac{M}{M_\odot} \right)^{1/2} \left(\frac{R}{10^9 \text{ cm}} \right)^{-1/2} \text{ cm} \quad (1)$$

where M and R are the white-dwarf mass and radius. The fragment size l is comparable to the instantaneous polar-cap size of MCVs. The emission will be in the hard X-ray regime if the accreting matter is optically thin to Thomson scattering; this is indeed observed in MCVs, which are strong hard X-ray emitters. One also requires that the density of the infalling material does not exceed a critical value, of the order of $10^{-6} \text{ g cm}^{-3}$, so that the kinetic energy of the infalling matter is small enough that it can be stopped above the photosphere¹⁴. This constraint is easily satisfied for typical dimensions of the

accreting blob as given by equation (1). The requirement that the accreting blob of matter be optically thin to Thomson scattering allows an estimate of its mass $\Delta M = \pi l^2 / \kappa_{\text{th}}$, where $\kappa_{\text{th}} = 0.19(1+X) \text{ cm}^2 \text{ g}^{-1}$ is the opacity, X being the mass abundance of hydrogen. In the case of water, which is thought to be the major constituent of comets, $\kappa_{\text{th}} = 0.21$. Then one finds

$$\Delta M = 1.3 \times 10^{17} \left(\frac{\Delta t}{0.1 \text{ s}} \right)^2 \left(\frac{M}{M_\odot} \right) \left(\frac{R}{10^9 \text{ cm}} \right)^{-1} \text{ g} \quad (2)$$

indicating a distance d of

$$d = 14 \left(\frac{F}{10^{-6} \text{ erg cm}^{-2}} \right)^{-1/2} \left(\frac{\Delta t}{0.1 \text{ s}} \right) \left(\frac{M}{M_\odot} \right)^{1/2} \left(\frac{R}{10^9 \text{ cm}} \right)^{-1/2} \text{ pc} \quad (3)$$

where F is the burst fluence. The accretion rate is thus a few times 10^{17} g s^{-1} , which is identical to that observed in MCVs, so that the accretion regime is much the same as in these systems.

The mean number of expected sources per year, τ^{-1} , can be estimated by rescaling the results of Tremaine and Zytow⁵ to the case of a white dwarf accreting a total cometary mass of 10^{17} g , which gives a value of $(250 \text{ yr})^{-1}$. The mass distribution of magnetic white dwarfs is not very well known; they are probably more massive than non-magnetic ones, as their progenitors are also more massive⁷. Assuming a mass of $1 M_\odot$, one obtains

$$\tau^{-1} = 0.13 \left(\frac{F}{10^{-6} \text{ erg cm}^{-2}} \right)^{-3/2} \left(\frac{\Delta t}{0.1 \text{ s}} \right)^3 \text{ yr}^{-1} \quad (4)$$

This result is uncertain by more than an order of magnitude. The main sources of errors are the determination of the recurrence rate, uncertainties in the mass of magnetic white dwarfs, and geometrical factors that enter equation (2) in the case of non-spherically symmetric geometries. Equation (4) predicts the detection of one source every 8 yr, whereas at least two (excluding the 5 March event, see below) have been observed in the past 15 yr. In view of the uncertainties above, this agreement is satisfactory. Assuming a source distance of about 15 pc, the visual magnitude is then in the range⁶ 11–15, which is too faint to be immediately identified in the 430-arcmin² error box of SGR 1806–21.

In this model, the emitted spectrum is an optically thin bremsstrahlung, with a temperature T_s of the order of the virial temperature; more precisely^{15,16},

$$kT_s = 52 \mu \left(\frac{M}{M_\odot} \right) \left(\frac{R}{10^9 \text{ cm}} \right)^{-1} \text{ keV} \quad (5)$$

where μ is the mean molecular weight per particle ($\mu = 1.38$ for water). The result of equation (5) is consistent with the observed spectrum ($kT = 40 \text{ keV}$) if the white dwarf has a mass of only $0.6 M_\odot$. The formula is relatively accurate for MCVs¹⁷ in which the mass of the white dwarf can be measured, for which values of $\mu = 0.615$ (corresponding to cosmic abundances) give $T_s = 15$ – 30 keV , as observed.

When the infalling blob hits the surface of the white dwarf, a shock forms at the leading edge and propagates backwards until the post-shock plasma can cool. At this stage a quasi-steady state is reached and the infalling matter crosses a stationary shock. The rise time of the burst, t_r , is therefore equal to the time required to form a stationary shock:

$$t_r = 7.6 \times 10^{-3} \left(\frac{M}{M_\odot} \right)^{1/2} \left(\frac{R}{10^9 \text{ cm}} \right)^{-1/2} \left(\frac{\rho}{10^{-6} \text{ g cm}^{-3}} \right)^{-1} \text{ s} \quad (6)$$

where ρ is the density of the shocked material at the surface of the white dwarf; this gives the correct result for the values of the density required for the emission to be in the hard X-ray regime¹⁴, typically $\rho \approx 10^{-6}$ – $10^{-7} \text{ g cm}^{-3}$.

The repeater GB790305 (the ‘5 March event’) had a rise time shorter than 0.1 ms and a significant harder spectrum (with, in

particular, a line feature at 400 keV), and therefore is difficult to reconcile with our model. If the association with the supernova remnant in the Large Magellanic Cloud is real, cometary accretion is in any case ruled out. On the other hand, if the association is coincidental, our model can explain all but the first burst, which would require extreme values of the parameters (for example, the mass of the infalling object). This burster is probably exceptional.

We predict, therefore, that there should be magnetic white dwarfs within the error boxes of SGRs. If their sizes correspond to a few arcmin², they should be detectable. In addition, because the cometary material is stopped above the white-dwarf photosphere (note that this need not be true in the case of non-magnetic white dwarfs), the stellar surface could be contaminated by cometary matter, in which case the white-dwarf spectrum would show abnormal lines¹⁰. A significant fraction (20–30%) of the energy of the hard X-rays can be absorbed by the white-dwarf surface and re-emitted in the X-ray ultraviolet, so a simultaneous burst at lower energy is expected and, because of the proximity of the sources, could be detectable. If, as in the case of MCVs, the accretion flow were to be unstable and several blobs were to form, some of them could penetrate below the surface and increase the X-ray ultraviolet flux. Optical bursts (or infrared, depending on the magnetic field strength at the white-dwarf surface) could also be produced by cyclotron radiation.

We thank Bohdan Paczyński for useful comments.

Received 16 September 1988; accepted 13 January 1989.

- Atteia, J. L. *et al. Astrophys. J.* **320**, L105–L110 (1987).
- Laros, J. G. *et al. Astrophys. J.* **320**, L111–L115 (1987).
- Hameury, J. M. & Lasota, J. P. in *Gamma-ray bursts* (eds Liang, E. P. & Petrosian, V.) 164–206 (Am. Inst. Phys., New York, 1986).
- Livio, M. & Taam, R. E. *Nature* **327**, 398–400 (1987).
- Tremaine, S. & Zytow, A. N. *Astrophys. J.* **301**, 155–163 (1986).
- Allen, C. W. in *Astrophysical Quantities* (Athlon Press, London, 1973).
- Angel, J. R. P., Borra, E. F. & Landstreet, J. D. *Astrophys. J. Suppl. Ser.* **45**, 457–474 (1981).
- Katz, J. I. *Astrophys. J.* **309**, 253–259 (1986).
- Hills, J. G. *Astrophys. J.* **267**, 322–333 (1983).
- Alcock, C., Friestrom, C. C. & Siegelman, R. *Astrophys. J.* **302**, 462–476 (1986).
- Harwit, M. & Salpeter, E. R. *Astrophys. J.* **186**, L37–L40 (1973).
- Burnard, D. J., Lea, S. M. & Arons, J. *Astrophys. J.* **266**, 175–187 (1983).
- Hameury, J. M., King, A. R. & Lasota, J. P. *Mon. Not. R. astr. Soc.* **218**, 695–710 (1986).
- Frank, J., King, A. R. & Lasota, J. P. *Astr. Astrophys.* **193**, 113–118 (1988).
- Hoshi, R. *Prog. theor. Phys.* **49**, 776–789 (1973).
- Aizu, K. *Prog. theor. Phys.* **49**, 1184–1194 (1973).
- Mukai, K. & Charles, P. A. *Mon. Not. R. astr. Soc.* **226**, 209–235 (1987).

Production of technetium in red giants by γ -ray-induced fission

Robert A. Malaney

Institute of Geophysics and Planetary Physics,
University of California, Lawrence Livermore National Laboratory,
California 94550, USA

The discovery of technetium in red giant stars¹ was a landmark in stellar nucleosynthesis theory. Because of the short half-lives of its isotopes, the presence of technetium has been regarded as irrefutable evidence for a recent episode of slow-neutron capture ('s-process') in asymptotic giant branch stars^{2–6}. But the 100 or so stars with technetium lines in their spectra^{7,8} (Tc stars) show the opposite correlation between metallicity and technetium production to that expected to result^{3,4,9} from an s-process. Here we argue that the production of technetium in red giant stars is not necessarily related to an interior s-process mechanism, and that some observations of technetium in metal-rich (population I) red giants may be more readily explained by γ -ray-induced fission of heavy isotopes (atomic mass $A \approx 240$) initially present in the stellar envelope.

The independent use of γ -rays and of spontaneous fission of heavy isotopes as a means of producing rare isotopes is not new^{10–16}. Here we show that observable quantities of technetium can be produced by a combination of such mechanisms, that is, by photofission. Both ingredients required for a photofission process, γ -rays and fissionable isotopes, are present in a stellar interior. The temperature of the hydrogen-burning shell during the quiescent period (that is, between thermal pulses) of low-mass asymptotic-giant-branch (AGB) evolution is typically $\sim 50 \times 10^6$ K. At such temperatures the CN cycle dominates and the main sources of energetic γ -rays are the reactions $^{13}\text{C} + \text{p} \rightarrow ^{14}\text{N} + \gamma$ and $^{14}\text{N} + \text{p} \rightarrow ^{15}\text{O} + \gamma$, which give rise to 7.55-MeV and 7.293-MeV γ -rays, respectively. Thus, for every four protons in the hydrogen shell that are converted into a ^4He nucleus, two γ -rays with energies in excess of 7 MeV are produced. Isotopes with $A \approx 240$ are known to exist both in the Solar System¹⁷ and in the atmospheres of stars¹⁸, and the fission threshold energies¹⁹ of such heavy nuclei (~ 6 MeV) are below the energy of these γ -rays produced in the CN cycle.

The heavy isotopes that we will consider here are ^{232}Th , ^{235}U and ^{238}U . The fission-threshold data¹⁹ of these heavy isotopes are well described by Seaborg's (see ref. 20) semi-empirical formula

$$T = 19.0 - 0.36 \frac{Z^2}{A} \text{ MeV} \quad (1)$$

where T is the fission threshold and Z the atomic number. According to equation (1) the fission threshold energies of the isotopes above are below the energies of the abundant CN γ -rays. To calculate the rate at which these nuclei undergo fission, we must estimate the fraction of the γ -rays that interact with a heavy isotope before the γ -ray energy is reduced below the fission threshold by Compton scattering. Because it varies only slowly with energy, we assume that the Compton-scattering cross-section $\sigma_{\gamma,e}$ may be approximated by the Klein–Nishina²⁰ formula,

$$\sigma_{\gamma,e} = \frac{3}{4} \sigma_T \left\{ \frac{1+\alpha}{\alpha^2} \left[\frac{2(1+\alpha)}{(1+2\alpha)} - \frac{1}{\alpha} \ln(1+2\alpha) \right] + \frac{1}{2\alpha} \ln(1+2\alpha) - \frac{(1+3\alpha)}{(1+2\alpha)^2} \right\} \quad (2)$$

where α is the ratio of the initial γ -ray energy, E , to the electron rest-mass, and $\sigma_T = 6.65 \times 10^{-25} \text{ cm}^2$. For $E = 7$ MeV, equation (2) gives $\sigma_{\gamma,e} \approx 60$ millibarn (mb; $1 \text{ mb} = 10^{-27} \text{ cm}^2$). The new energy E' of the γ -ray after a single collision with an electron can be estimated as²⁰ $E'/E = [1 + \alpha(1 - \cos \phi)]^{-1}$, where ϕ is the scattering angle in the rest frame of the electron. A single scattering event can therefore reduce the energy of a γ -ray below the fission threshold. As a consequence, we will make the simplifying assumption that if a γ -ray is scattered by an electron before it interacts with a heavy nucleus then it no longer plays any role in the photofission process.

From a consideration of the CN cycle operating at a temperature of 50×10^6 K in a population I AGB star, one can show that the number of high-energy γ -rays (> 7 MeV) produced in the hydrogen shell is $n_\gamma \approx 10^{14} \text{ cm}^{-3} \text{ s}^{-1}$. The fraction of these γ -rays absorbed by a fission nucleus without prior Compton scattering is given by

$$f = \frac{\sigma_{\gamma,f} N_f}{\sigma_{\gamma,e} N_e} \quad (3)$$

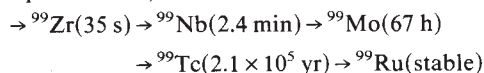
where $\sigma_{\gamma,f}$ is the photofission cross-section and N_f and N_e are the heavy-isotope number density (the sum of ^{232}Th , ^{235}U and ^{238}U) and the electron number density, respectively. For isotopes with $A \approx 240$ the photofission cross-section can be written as²¹

$$\sigma_{\gamma,f}(E) \approx 52(E/7 \text{ MeV})^3 (\Gamma_f/\Gamma_T)_E \text{ mb} \quad (4)$$

where $\Gamma_T = \Gamma_f + \Gamma_n + \Gamma_\gamma$ is the sum of all possible decay widths (Γ_n is the neutron decay width). This simple formula agrees well

with the available photofission-cross-section data¹⁹. Generally, for a ~ 7 -MeV excitation in a heavy isotope, $(\Gamma_f/\Gamma_\gamma) \approx 1$ and $\Gamma_f > \Gamma_\gamma$. De-excitation by γ -ray emission may be neglected for our purposes¹⁹. Any heavy isotope produced in a (γ, n) reaction, such as ^{231}Pa and ^{237}Np , is also susceptible to photofission (in fact, the (γ, n) thresholds of these latter isotopes are higher, ~ 7 MeV, so the fission branching ratios will also be higher). We will therefore assume that fission is the most probable end result of absorption of an energetic γ -ray by thorium or uranium. Equation (4) indicates that, for the three heavy isotopes considered here, the photofission cross-section for a 7-MeV γ -ray is comparable to the Compton-scattering cross-section. From equations (2), (3) and (4), it follows that a fraction $\sim (1 + X_H)^{-1}$ of all the heavy isotopes will interact with a γ -ray that has not been Compton-degraded (here, X_H is the hydrogen mass fraction). Thus, even in this simple model in which we disregard all scattered photons, a large fraction of the heavy isotopes will interact with a high-energy γ -ray. To obtain an order-of-magnitude estimate, we will assume that all of the heavy isotopes undergo γ -ray-induced fission.

To determine the technetium production rate, we must estimate the fission yields of technetium from the heavy isotopes. Fission of isotopes with $A \approx 240$ is generally characterized by asymmetric fission yields with mass peaks at $A \approx 100$ and $A \approx 140$. Only the fission production of the neutron-rich isotope ^{99}Zr leads to a relatively long-lived isotope of technetium. The radioactive chain leading to the production of ^{99}Tc is (half-lives shown in parentheses)



The yield P of the photofission fragments that lead to the formation of ^{99}Tc is generally in the range 2–7% (ref. 19). We assume $P = 0.05$ for the total yield of technetium as a result of γ -ray-induced fission reactions. The number density of technetium produced in the hydrogen shell is then $N_{\text{Tc}} = PN_f$.

Following a helium-shell flash on the AGB, the hydrogen-burning shell is pushed outwards towards cooler regions of the star, and at this point the envelope convection extends well below the hydrogen-processed material^{2–4}. This is known to occur for both population I and population II AGB models, and corresponds to extension of the convection envelope into the carbon- and s-process-enriched intershell region which is found only in low-metallicity models^{4,9}. Thus any technetium produced in the hydrogen shell during its previous burning phase will be swept into the stellar atmosphere, where it can then be observed. This contrasts with the case of envelope convection on the first red-giant branch. In that case, convection never extends as far down as the hydrogen-burning shell²², so any technetium produced in the hydrogen shell on the first red-giant branch is never transported to the stellar envelope.

To compare these calculations with observation we must estimate the dilution factor $d = n_p \Delta M_{\text{sh}} / M_{\text{env}}$, where n_p is the number of thermal pulses undergone by the star, ΔM_{sh} is the mass processed through the hydrogen-burning shell during one interpulse period, and M_{env} is the mass of the stellar envelope. The estimate of the dilution factor is the most uncertain parameter. For low-mass AGB models, values of $\Delta M_{\text{sh}} \approx 0.01 M_\odot$ and $n_p \approx 10$ are typical^{3,23}. The envelope masses of the Tc stars are not known at present. The situation is complicated by mass loss on the AGB and its subsequent effect on the stellar evolution^{4,23}. However, low-mass AGB models that assume envelope masses in the range $0.05 M_\odot \leq M_{\text{env}} \leq 1.0 M_\odot$ give good agreement with the observed general characteristics of the Tc stars^{2–4}. Our definition of the dilution factor is somewhat ambiguous because the envelope mass is time-dependent. But from the values of the relevant stellar parameters indicated above, it would appear that dilution factors in the range $0.1 < d < 1$ are not unreasonable ($d \approx 1$ is approached as mass loss reduces the envelope mass to $M_{\text{env}} \approx \Delta M_{\text{sh}}$).

Using our simple approximations, the production of technetium in the atmospheres of red giant stars is given by

$$\frac{dN_{\text{Tc}}}{dt} \approx dPN_f - \lambda N_{\text{Tc}} \quad (5)$$

where λ is the beta decay rate of the technetium isotope. For timescales of $\sim 10^5$ years, with d in the range indicated above and with the initial heavy-isotope abundance in the range 0.5–4 times that of the Solar System¹⁷ (the abundances of Th and U predicted by rapid neutron capture (the r-process) are model-dependent), equation (5) predicts an observed technetium abundance in the range 10^{-2} – 10^{-5} (relative to a Si abundance of 10^6) for the population I AGB model. Of the 100 or so Tc stars presently known to exist, the abundance of technetium in the atmospheres of only fifteen have been actually determined^{24–27}. At least 50% of these stars have technetium abundances comparable with, or lower than, the upper limit of technetium production estimated above. However, the high abundances of technetium (0.1–1 on the Si scale above) observed in a few Tc stars, such as χ Cyg and R And^{26,27}, require a more efficient means of producing technetium in addition to the photofission scheme outlined here.

For those Tc stars that possessed higher or lower initial heavy-isotope abundances, the rate of technetium production would be altered accordingly. In particular, the production of technetium by the photofission mechanism in population II stars would be greatly reduced simply as a consequence of the lower availability of seed nuclei. For a population II star possessing $\sim 1/20$ of the population I initial heavy-isotope abundances, the production of technetium in the stellar atmosphere would probably be at such a low level that it would be very difficult to detect (a technetium abundance of $\sim 10^{-3}$ on the Si scale is the limit of detectability of the most recent survey)⁸. Furthermore, if the population II stars are sufficiently old, then the radioactive decay of their initial thorium and uranium may also be a factor that reduces technetium production. In contrast, the s-process production of technetium, at a level consistent with other s-process elements, should be detectable in a population II star for several ^{99}Tc half-lives.

In summary, we propose that the technetium seen in the atmospheres of some population I AGB stars is not necessarily a consequence of s-process production in the stellar interior, as hitherto believed, but may be simply a result of γ -ray-induced fission of heavy nuclei present in the hydrogen-burning shell. Our calculation of the technetium production can provide only an order-of-magnitude estimate because of uncertainties in the initial heavy-element abundances and in the dilution factors. But this does not alter the qualitative conclusion that the detection of technetium in a star cannot be regarded as a proof of a recent s-process episode. We are now carrying out a detailed study of low-mass AGB evolution and photofission processes, to further quantify the basic conclusions outlined here. Finally, we note that the destruction of the actinides, as described here, has wider implications, specifically with regard to the Th/Nd ratio and other nuclear chronometers²⁸.

I acknowledge discussions with G. Bazan, D. Lambert, J. Lattanzio, G. Matthews, B. Meyer and G. Wallerstein. This work was performed under the auspices of the US Department of Energy by the Lawrence Livermore National Laboratory.

Received 10 November 1988; accepted 11 January 1989.

- Merrill, P. *Astrophys. J.* **116**, 21–26 (1952).
- Iben, I. Jr & Renzini, A. *A. Rev. Astr. Astrophys.* **21**, 271–342 (1983).
- Hollowell, D. E. thesis, Univ. of Illinois (1988).
- Lattanzio, J. C. in *Proc. Evolution of Peculiar Red Giant Stars*, IAU Coll. **106**, (Indiana Univ.) (1988).
- Mathews, G. J., Takahashi, K., Ward, R. A. & Howard, W. M. *Astrophys. J.* **302**, 410–414 (1986).
- Malaney, R. A. *Astrophys. J.* **321**, 832–845 (1987).
- Little, S. J., Little-Marengo, I. R. & Hagen Bauer, W. *Astr. J.* **94**, 981–995 (1987).
- Smith, V. V. & Lambert, D. L. *Astrophys. J.* **333**, 219–226 (1988).
- Iben, I. Jr & Renzini, A. *Astrophys. J.* **259**, L79–L83 (1983).
- Harrison, T. G. *Astrophys. Lett.* **18**, 7–12 (1976).
- Harrison, T. G. *Astrophys. Lett.* **16**, 87–95 (1975).

12. Dakowski, M. *Earth planet. Sci. Lett.* **6**, 152 (1969).
13. Kuchowicz, B. *Astrophys. Lett.* **9**, 85–89 (1971).
14. Kuroda, P. K. *Nature* **187**, 36–38 (1960).
15. Alexander, E. C., Lewis, R. S., Reynolds, J. H. & Michel, M. D. *Science* **172**, 837–840 (1971).
16. Selinov, I. P. in *Proc. 2nd UN Int. Conf. on the Peaceful uses of Atomic Energy* **30**, 307–326 (Unit. Nations, Geneva, 1958).
17. Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2362–2380 (1982).
18. Cowley, C. R. & Adelman, S. J. *Astrophys. Lett.* **16**, 5–7 (1975).
19. Gorbachev, V. M., Zamyatnin, Y. S. & Lbov, A. A. *Nuclear Reactions in Heavy Elements* (Pergamon, Oxford, 1980).
20. Evans, R. D. *The Atomic Nucleus* (McGraw-Hill, New York, 1955).
21. Vandebosch, R. & Huizenga, J. R. *Nuclear Fission* (Academic, New York, 1973).
22. Sweigart, A. V. & Gross, P. G. *Astrophys. J. Suppl.* **36**, 405–437 (1978).
23. Boothroyd, A. I. & Sackmann, I.-J. *Astrophys. J.* **238**, 671–679 (1988).
24. Smith, V. V. in *Origin and Distribution of the Elements* (ed. Mathews, G. J.) 535–541 (World Scientific, Singapore, 1988).
25. Smith, V. V. & Lambert, D. L. *Astrophys. J.* **311**, 843–863 (1986).
26. Dominy, J. F. & Wallerstein, G. *Astrophys. J.* **310**, 371–377 (1986).
27. Wallerstein, G. & Dominy, J. F. *Astrophys. J.* **330**, 937–941 (1988).
28. Malaney, R. A., Mathews, G. J. & Dearborn, D. S. P. *Astrophys. J.* (submitted).

Nature of the charge carriers in electron-doped copper oxide superconductors

J. M. Tranquada*, S. M. Heald*, A. R. Moodenbaugh*, G. Liang† & M. Croft†

* Brookhaven National Laboratory, Upton, New York 11973, USA

† Physics Department, Rutgers University, Piscataway, New Jersey 08855-0849, USA

The charge carriers in all of the well-known layered copper oxide high-temperature superconductors^{1–4} are holes located in the CuO₂ planes, the common structural element of these systems. Recently, Tokura *et al.*⁵ have reported that doping of Ln₂CuO₄, where Ln represents Pr, Nd or Sm, with formally tetravalent Ce can produce a high-temperature superconductor with electron carriers. To determine the character of the carrier states in Nd_{2-x}Ce_xCuO₄, we have performed X-ray absorption measurements at the Cu K and Ce L₃ edges for several different Ce concentrations. We find clear evidence that the electrons introduced by doping fill 3d holes on the Cu atoms, thus converting some Cu²⁺ ions to Cu⁺. Hence the nature of the states nearest to the Fermi energy (that is, those responsible for conduction) is qualitatively different from that in the hole-doped superconductors, in which the holes move in a band formed predominantly from O 2p states^{8–10}. As a result, the details of the electron-pairing mechanism could be very different in the two systems¹¹.

In the prototypical hole-doped superconductor system La_{2-x}Sr_xCuO₄, each Cu atom is surrounded by a tetragonally distorted octahedron of O atoms. In the undoped compound, each Cu ion is in a 3d⁹ configuration (2+ valence), whereas the 2p states of the oxygen atoms are fully occupied. When Sr²⁺ is substituted for La³⁺, the resulting holes enter the O 2p band; the copper atoms retain their 2+ valence because of the large intra-atomic Coulomb energy^{6–10}. As pointed out by Emery¹¹, the electrostatic potential from the two out-of-plane oxygen neighbours inhibits fluctuations of electrons onto a Cu site, but it helps to stabilize the presence of holes in the plane. In Nd_{2-x}Ce_xCuO₄, the out-of-plane O atoms shift away from the Cu sites⁵, thus making the potential more favourable for addition of an electron to fill the Cu 3d hole. This argument is consistent with the empirical observation that Nd₂CuO₄ can be doped only with electrons⁶, whereas La₂CuO₄ will accept only holes.

To understand the pairing mechanism(s) in copper oxide superconductors, it is important to establish whether the added electrons in Nd_{2-x}Ce_xCuO₄ are correlated, resulting in mixed-valent copper, or whether they are distributed in a band such that each copper sees an average increase in charge. A secondary question concerns the behaviour of Ce. In CeO₂, an insulator used as a starting material for making doped samples, band-structure calculations¹² and a many-body analysis^{13,14} of core-level photoemission and X-ray absorption measurements indicate that Ce has an intermediate valence of approximately 3.5+,

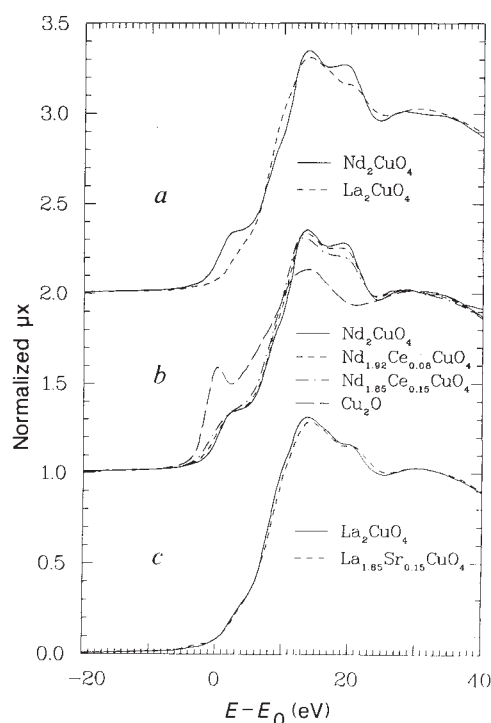


Fig. 1 X-ray absorption near-edge structure (plotted as X-ray absorption, μx , versus energy) measured at the Cu K edge ($E_0 = 8,980$ eV) in a series of copper oxide compounds. The data have been normalized to an edge jump $\Delta\mu x = 1$, and some of the spectra have been shifted vertically for clarity.

rather than 4+, because of hybridization with the O 2p band. To understand the doping process it is important to determine the valence of the substituted Ce.

To answer these questions we performed X-ray absorption measurements at the Cu K and Ce L₃ edges in Nd_{2-x}Ce_xCuO₄ with $x = 0, 0.075, 0.15$ and 0.2 . Oxidized samples were prepared essentially by the recipe given in ref. 5. X-ray diffraction measurements indicated that each sample had the T'-phase structure^{5,15}, with no detectable impurities. A superconducting sample with an onset temperature of 20 K, determined by a.c. susceptibility measurements, was prepared by annealing a sample with $x = 0.15$ in Ar at 900 °C. The X-ray absorption measurements were performed at beam line X-11 at the National Synchrotron Light Source, Brookhaven National Laboratory, using a water-cooled Si (111) double-crystal monochromator. Each sample was ground to a fine powder, rubbed onto Scotch tape, and measured in transmission at room temperature. The energy was calibrated by simultaneous measurement of a reference material.

In Fig. 1a we compare the X-ray absorption near-edge spectra observed at the Cu K edge for Nd₂CuO₄ and La₂CuO₄. The spectrum for the latter compound has been discussed in detail elsewhere^{6,7}. The main peak at ~13 eV is due to a Cu 4p-like final state, and its energy is indicative of a 3d⁹ ground-state configuration. Crystal-field effects split the out-of-plane ($4p_z$) and in-plane ($4p_x$) states by ~4 eV, as demonstrated by polarization-dependent measurements on magnetically orientated samples¹⁶. The shoulder at ~3 eV is due to a $1s \rightarrow 4p_z$ transition combined with the transfer of an electron from an O neighbour into the Cu 3d hole (a 'shakedown' transition)^{17,18}. Note that the shakedown feature observed for Nd₂CuO₄ is much more prominent and at lower energy. This indicates that an electron added to the 3d hole of a copper atom is more tightly bound in the Nd compound than in the La one, as expected from the difference in structure.

How can we tell whether any Cu⁺ is created by doping Nd₂CuO₄ with Ce? The splitting of the 4p final states in the

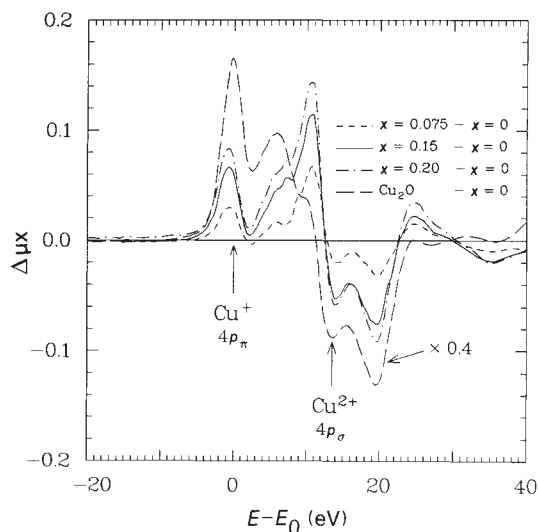


Fig. 2 Cu K edge difference spectra for $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$. The absorption spectrum for Nd_2CuO_4 has been subtracted from each of the other curves shown in Fig. 1b; a curve for $x = 0.2$ is also included. The positions of the $\text{Cu}^+ 4p_\pi$ and $\text{Cu}^{2+} 4p_\sigma$ features are indicated by arrows. The curve for Cu_2O has been scaled down by a factor of 0.4. The intensity changes caused by Ce doping, compared with the Cu_2O difference, are roughly consistent with one $\text{Cu}^{2+} \rightarrow \text{Cu}^+$ transformation for each Ce atom added.

presence of Cu^+ should be the same as that for Cu^{2+} , whereas the average $1s \rightarrow 4p$ transition energy should decrease by ~ 9 eV (refs 6, 7). The appearance of a Cu^+ feature should also be correlated with a decrease of the Cu^{2+} feature. Comparing the spectra of the (oxidized) Ce-doped samples with that of the undoped material in Fig. 1b, one can see that with increasing Ce content the Cu^{2+} peak decreases and extra intensity appears in the region corresponding to the $\text{Cu}^+ 4p_\pi$ feature that is prominent in Cu_2O . (The trend is especially clear in the difference spectra of Fig. 2.) This is precisely the behaviour expected for a $\text{Cu}^+/\text{Cu}^{2+}$ mixture. The changes are clearly not consistent with a simple edge shift, as might be expected for an uncorrelated single-band model. Surprisingly, the spectra of the oxidized and reduced $\text{Ce}_{0.15}$ samples were identical. Hence, the role of reduction, in which a fairly small amount of oxygen is removed, is not yet clear.

To emphasize the difference from the hole-doped materials, the effect of Sr doping in La_2CuO_4 is illustrated in Fig. 1c. There is a very slight shift in the edge to higher energy, which is at least partially expected because of the decrease in Cu-O spacing, but there are no qualitative changes in the edge features. Measurements at the oxygen K edge⁸ in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4-y}$ and $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ have shown clearly that the main effect of doping is to introduce holes into the O 2p band. Emery¹¹ has pointed out that the presence of Cu^+ ions in the CuO_2 planes of $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ could induce holes in O 2p levels through inter-

atomic Coulomb interactions; however, an electron leaving an oxygen atom would have to go onto a Cu^{2+} ion, thus enhancing the Cu^+ population.

Figure 3 shows absorption spectra measured at the Ce L_3 edge in CeO_2 and $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_4$. The valency is related to the relative magnitude of the two near-edge peaks, and it seems that the average valence of Ce is essentially the same in the two compounds. Hence, the average valence of Ce substituted into Nd_2CuO_4 is $\sim 3.5+$ rather than the formal valence of $4+$. However, if this reduced valence of Ce is compensated by holes on neighbouring oxygens arising from hybridization, the net doping on Cu sites could be similar to that expected for Ce^{4+} . We have also measured the Nd L_2 edge in the Nd compounds, and we find a single sharp peak at the edge which shows essentially no change with doping and which looks nearly identical to that found in Nd_2O_3 .

To conclude, X-ray absorption measurements at the Cu K edge in $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ show that the electrons introduced into the CuO_2 planes by Ce doping fill Cu 3d holes, resulting in a highly correlated mixture of Cu^+ and Cu^{2+} ions. The results are inconsistent with a single-band description. The Ce dopants have an intermediate valence which is significantly less than the formal value of $4+$. Measurements at the oxygen K edge would provide useful further information. In any case, it is clear that the nature of the charge carriers in the electron-doped superconductors is quite different from that in the hole-doped materials. These observations should put a significant constraint on theories of electron-pairing that attempt to describe simultaneously all of the copper oxide superconductors. It seems quite possible that the details of electron-pairing may be different in the electron- and hole-doped systems.

We are grateful to V. J. Emery for valuable discussions. This work was supported by the US Department of Energy.

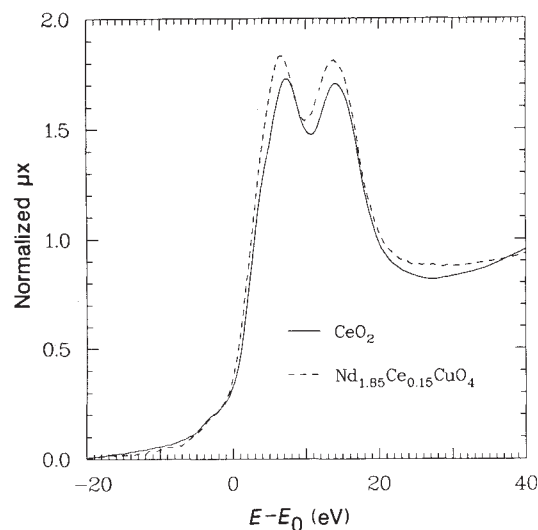


Fig. 3 Ce L_3 near-edge structure measured in CeO_2 and $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_4$. $E_0 = 5,723$ eV.

Received 2 February; accepted 8 February 1989.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189-194 (1986).
2. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908-910 (1987).
3. Maeda, H., Tanaka, Y., Fukutomi, M. & Asano, T. *Jap. J. appl. Phys.* **27**, L209-L210 (1988).
4. Sheng, Z. Z. & Hermann, A. M. *Nature* **332**, 55-58 (1988).
5. Tokura, Y., Takagi, H. & Uchida, S. *Nature* **337**, 345-347 (1989).
6. Tranquada, J. M., Heald, S. M., Moodenbaugh, A. R. & Suenaga, M. *Phys. Rev. B* **35**, 7187-7190 (1987).
7. Tranquada, J. M., Heald, S. M. & Moodenbaugh, A. R. *Phys. Rev. B* **36**, 5263-5274 (1987).
8. Yarnoff, J. A. *et al. Phys. Rev. B* **36**, 3967-3970 (1987).
9. Nücker, N. *et al. Phys. Rev. B* **37**, 5158-5163 (1988).
10. Emery, V. J. *Phys. Rev. Lett.* **58**, 2794-2797 (1987).
11. Emery, V. J. *Nature* **337**, 306 (1989).
12. Koelling, D. D., Boring, A. M. & Wood, J. H. *Solid St. Commun.* **47**, 227-232 (1983).
13. Kotani, A., Mizuta, H., Jo, T. & Pärlebas, J. C. *Solid St. Commun.* **53**, 805-810 (1985).
14. Jo, T. & Kotani, A. *Solid St. Commun.* **54**, 451-456 (1985).
15. Müller-Buschbaum, Hk. *Angew. Chem. Int. Ed. Engl.* **16**, 674-687 (1977).
16. Heald, S. M., Tranquada, J. M., Moodenbaugh, A. R. & Xu, Y. *Phys. Rev. B* **38**, 761-764 (1988).
17. Bair, R. A. & Goddard III, W. A. *Phys. Rev. B* **22**, 2767-2776 (1980).
18. Kosugi, N., Yokoyama, T., Asakura, K. & Kuroda, H. *Chem. Phys.* **91**, 249-256 (1984).

Use of trialkylamines as an indicator of urban sewage in sludges, coastal waters and sediments

M. Valls, J. M. Bayona & J. Albaigés

Environmental Chemistry Department, Centro de Investigación y Desarrollo (CSIC), Jordi Girona, 18-26, 08034-Barcelona, Spain

Coastal areas receive a variety of land-based organic inputs, both natural and synthetic. The identification of specific organic markers in different marine compartments may be used to trace the pathways of transport, the regions of concentration and the short-term effects of such inputs into the ocean. Particular attention is being devoted to the assessment of urban wastes at coastal areas because of the environmental implications of discharging large quantities of urban sewage waters or sludges into the sea. Here we report the widespread occurrence of trialkylamines (TAMs) in coastal areas, originating from such sewage discharges. The concentrations of these compounds—which have the formula $R_1R_2NCH_3$, where $R_1 = CH_3$ or C_nH_{2n+1} and $R_2 = C_nH_{2n+1}$, for $n = 14-18$ —in Western Mediterranean and North Sea coastal sediments (Fig. 1) ranged from 0.1 to 35 $\mu g\ g^{-1}$, depending on the distance from and the pollutant budget of the urban sewage sources. We have identified TAMs as trace impurities in quaternary ammonium salts used as fabric softeners in household laundry detergents, for which reason they are found in urban sewage waters, primarily associated with the particulate phase. They correlate in the sedimentary record with other domestic surfactant markers, such as the long-chain alkylbenzenes (LABs). We propose, therefore, that TAMs may serve as conservative indicators of urban sewage contamination of coastal areas.

A comprehensive review of the available potential markers of sewage sludges in marine sediments has already been pub-

lished¹. Among the different categories of tracers, certain groups of organic compounds and individual components have been considered, including silicones², chlorinated hydrocarbons³, plasticizers⁴, PAHs⁵, coprostanol⁶ and α -tocopherol acetate⁷. Very recently, benzthiazoles have also been proposed as potential indicators of the contamination of estuarine sediments by urban street runoff⁸. But the persistence of these markers in the marine environment, or their unique relation with urban sources^{1,5,9,10}, has not been fully established, thus bringing into question their applicability in environmental pollution assessment and, in particular, their exploitation for quantitative purposes and in historical recording studies.

In view of the great variety of cleaners and other surface-active products currently used by household consumers, particular attention has been devoted to the identification in municipal sludges and aquatic sediments of some refractory products derived from domestic surfactants. Among these, alkylphenols¹¹ and linear alkylbenzenes^{12,13}, which are indicative, respectively, of non-ionic and anionic detergents, have been described. During a systematic investigation of mutagenic organic compounds in sediments off-shore from Barcelona¹⁴, Spain, we identified a new group of conservative and widely distributed pollutants, the trialkylamines (TAMs), which can be used as specific markers of cationic surfactants.

The organic extracts of freeze-dried coastal sediments and sludges (chloromethane:methanol; 2:1) and waters (chloromethane), after separation of the acid fraction, were resolved by column chromatography (using a column of 8 g 5% water-deactivated neutral alumina and silica 8 g¹⁵) into fractions of increasing polarity by elution with hexane (20 ml), chloromethane-methanol (95:5; 50 ml) and diethyl ether (50 ml). In the latter fraction we found from gas chromatography-flame ionization detection (GC-FID) a homologous series of components exhibiting an even-odd predominance of carbon number (Fig. 2b), similar to that observed using a nitrogen-

Table 1 Concentrations of TAMs, LABs and UCM of hydrocarbons in urban sewage waters and sludges, coastal sea water and sediments

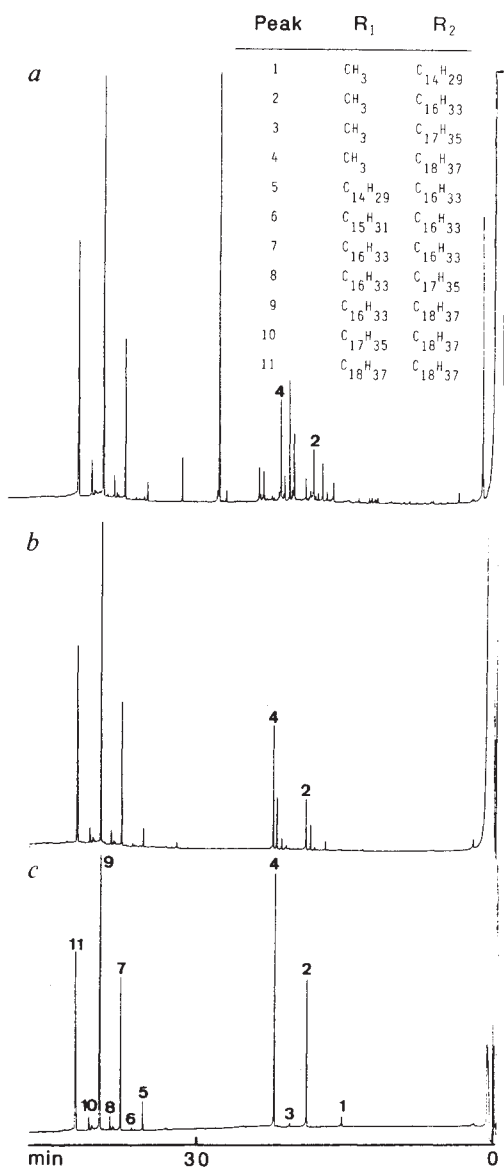
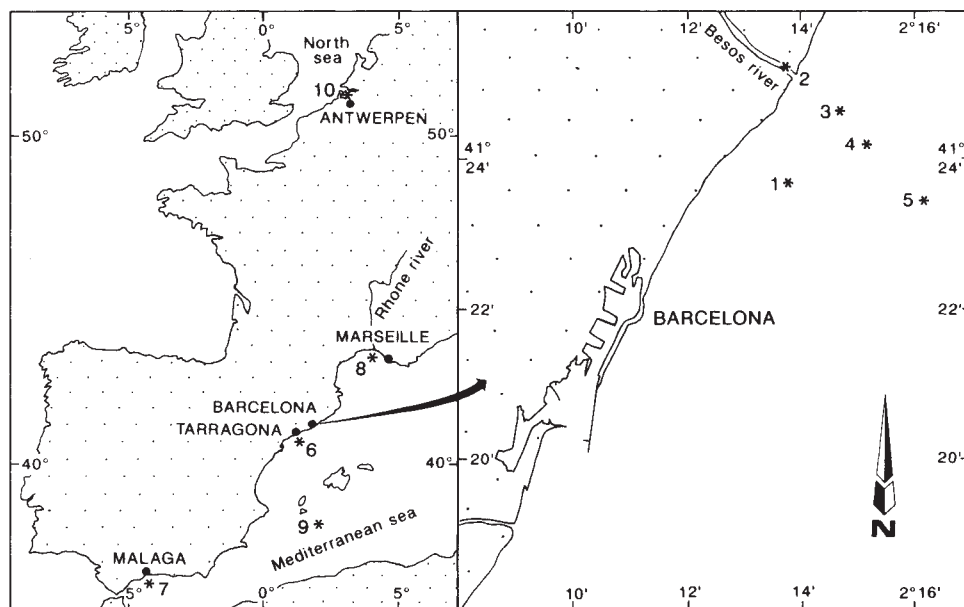
Sampling site*	Sample	Water depth (m)	TAM (total)	LAB (total)	UCM
Sewage waters† ($\mu g\ l^{-1}$)					
Barcelona	Dissolved (<0.45 μm)		189 $\pm$ 32	33 $\pm$ 12	60 $\pm$ 25
	Particulate		201 $\pm$ 45	41 $\pm$ 14	140 $\pm$ 42
Tarragona	Total		27 $\pm$ 5	—	19 $\pm$ 9
Sewage sludges ($\mu g\ g^{-1}$)					
München			380	—	—
Barcelona			26	—	130
Coastal water ($\mu g\ l^{-1}$)					
1	Dissolved (<0.45 μm)	25	0.005	<0.002	0.12
	Particulate	25	0.03	0.01	0.05
Sediments ($\mu g\ g^{-1}$)					
2	Besós estuary (0–5 cm)	2	35	27.5	168.7
3	Barcelona (offshore) (0–5 cm)	38	26	3.6	85.9
4	Barcelona (offshore) (0–2.5 cm)	50	6.1	1.31	26.4
	(2.5–5 cm)		4.6	0.49	25.1
	(5–7.5 cm)		0.5	0.29	17.8
	(8–12 cm)		0.1	0.11	4.4
	(12–17 cm)		ND	ND	0.7
	(17–22 cm)		ND	ND	0.2
	(22–27 cm)		ND	ND	ND
5	Barcelona (offshore) (0–2.5 cm)	65	2.3	0.72	12.4
6	Tarragona (offshore) (0–5 cm)	16	0.15	0.09	7.9
7	Malaga (offshore) (0–5 cm)	28	0.13	0.10	0.23
8	Rhone estuary (0–5 cm)	57	0.8	—	—
9	Balearic basin (0–2 cm)	1,400	ND	ND	0.05
10	Schelde estuary (0–5 cm)	15	4.9	—	—

ND: not detected; —: not determined.

* Numbers correspond to those indicated in Fig. 1.

† Mean of three determinations at different periods.

Fig. 1 Sediment sampling sites in the Western Mediterranean and North Sea.



selective detector (NPD) (Fig. 2c). The electron-impact (EI) mass spectra of this series of peaks (Fig. 3a shows the spectrum for peak 9) exhibited, respectively, a series of odd molecular ions ranging from 479 to 535 at intervals of 14 daltons. The positive-ion chemical-ionization (PCI) mass spectra displayed the $(M-1)^+$ ion as the base peak (Fig. 3b), in accordance with a hydride-ion-abstraction characteristic of an aliphatic molecule¹⁶. The base peaks in the EI mass spectra ($240+14n$, $n=0$ to 4) were also present in the PCI mass spectra, which were assigned to a β -cleavage of an alkylamine followed by a proton migration. The $m/z=58$ peak was assigned to the $((CH_3)_2NCH_2)^+$ ion (Fig. 3a). Confirmation that these compounds were TAMs was obtained by comparison of their EI mass spectra and retention times with those of N,N-dihexadecyl-N-methylamine, synthesized in the laboratory from N,N-dihexadecylamine.

N,N-dimethyl-N-alkylamines were also identified, but at lower concentrations (Fig. 2). However, the three long-chain (C_{14} – C_{18}) substituted alkylamines were not found either by high-temperature GC-NPD (320 °C) or by liquid chromatography-mass spectrometry using a thermospray interface (Hewlett-Packard 5988A).

To search for the source of these components in the environment, we investigated their presence in dimethylditallow-ammonium chloride (DMDTAC), a cationic surfactant used as a fabric softener in powdered household laundry detergents, which account for 10% of the total industrial surfactant output¹⁷. TAMs were found at concentrations of 450–500 $\mu\text{g g}^{-1}$ in this surfactant, and concentrations ranging from 2 to 500 $\mu\text{g l}^{-1}$ were found in different municipal sewage waters (Table 1). TAMs were usually more abundant in the aquatic environment than were LABs, another proposed urban-sewage marker^{10,12}, and would thus enable the distribution of such inputs in the sea to be traced at higher dilution levels. Moreover, TAMs apparently resist the biological treatment of these waste waters, because urban sewage sludges are particularly enriched in such compounds (Table 1). This may pose additional problems not only to their disposal into the sea but also in landfills.

Fig. 2 Representative high-resolution gas chromatography profiles of TAMs in a, particulate sea water (FID); b, sediment (FID); and c, sediment (NPD). Column is fused silica CP-5-SIL, 20 m $\times$ 0.25 mm inside diameter, programmed from 60 °C to 300 °C at 4 °C min^{-1} . Carrier gas is hydrogen at 50 cm s^{-1} .

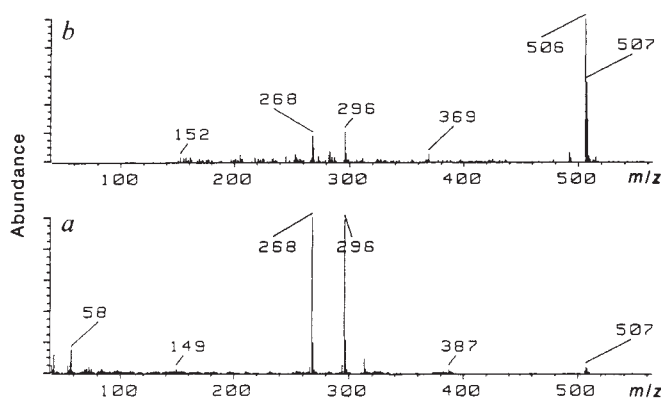


Fig. 3 Mass spectra of N-hexadecyl-N-methyl-N-octadecylamine (peak 9 in Fig. 2) obtained by a, EI at 70 eV, and b, PICI using 1.5 torr of methane in the ion source.

The distribution of TAMs in surface sediments from the Mediterranean and North Sea coasts exhibited maximum concentrations at the vicinity of urban areas, namely Barcelona, Marseille (Rhône estuary), Antwerpen (Schelde estuary), Tarragona and Malaga, and decreased with increasing distance from the source (Table 1 and Fig. 1), indicating inputs of domestic origin. The vertical profile obtained in a sediment sample offshore from Barcelona (Table 1) indicates a more recent input than that of the LABs, use of which was introduced in the mid-1960s in developed countries¹². In turn, both markers decreased more rapidly with depth relative to another anthropogenic indicator, the unresolved complex mixture of hydrocarbons (UCM) attributable to petrogenic sources.

A major sedimentary pathway for these pollutants could be the precipitation of particulate material in the water column. Although TAMs have been identified at similar high concentrations in both dissolved and particulate phases of the Barcelona sewage waters, they were found at much higher concentrations in particulates of coastal waters (Table 1). Such behaviour is very different to that of the UCM, probably because the two compounds have different sediment sorption partition coefficients (K_p). Therefore, adsorption and precipitation seems to be an effective mechanism in transporting TAMs from the water column towards the sediment.

The risk associated with the presence of TAMs in the marine environment is still unknown. Although aromatic amines are well-known carcinogens¹⁸, data on aliphatic amines are lacking. Investigation is now in progress to assess their accumulation in benthic biota and their ecotoxicity.

M.V. thanks the Ministry of Education (Spain) for a PhD fellowship. This is a contribution to the EEC Project EROS 2000 and to the Spanish MEDPOL Programme.

Received 14 December 1988; accepted 6 January 1989.

1. Vivian, C. M. G. *Sci. Tot. Envir.* **53**, 5-40 (1986).
2. Pellenberg, R. *Mar. Pollut. Bull.* **10**, 267-269 (1979).
3. McIntyre, A. E. & Lester, J. N. *Water Air Soil Pollut.* **23**, 397-415 (1984).
4. Giam, C. S., Chan, H. S., Neff, G. S. & Atlas, E. L. *Science* **199**, 419-421 (1978).
5. Gryzbowski, J., Radecki, A. & Rewkowska, G. *Envir. Sci. Technol.* **17**, 44-47 (1983).
6. Walker, R. W., Wun, Ch. K. & Litsky, W. *CRC Crit. Rev. Envir. Cont.* **12**, 91-112 (1982).
7. Eganhouse, R. P. & Kaplan, I. R. *Envir. Sci. Technol.* **19**, 282-285 (1985).
8. Spies, R. B., Andresen, B. D. & Rice, D. W. *Nature* **327**, 697-699 (1987).
9. Pocklington, R., Leonard, J. D. & Crewe, N. F. *Oceanol. Acta* **10**, 83-89 (1987).
10. Eganhouse, R. P., Olague, D. P., Gould, B. R. & Phinney, C. S. *Mar. Envir. Res.* **25**, 1-22 (1988).
11. Giger, W. E., Stephenou, E. & Schaffner, C. *Chemosphere* **10**, 1253-1263 (1981).
12. Ishiwatari, R., Takada, H., Yun, S. & Matsumoto, E. *Nature* **301**, 599-600 (1983).
13. Eganhouse, R. P., Blumfield, D. L. & Kaplan, I. R. *Envir. Sci. Technol.* **17**, 523-530 (1983).
14. Grifoll, M. et al. *Toxicity Assessment* **3**, 315-330 (1988).
15. Aceves, M. et al. *J. Chromatogr.* **436**, 503-509 (1988).
16. Harrison, A. G. in *Chemical Ionization Mass Spectrometry* 12-21 (CRC Press, Lausanne, 1982).
17. Greek, B. F. *Chem. Engng News* Jan. 25, 21-53 (1988).
18. Garner, R. C., Martin, C. N. & Clayton, D. B. in *Chemical Carcinogens*, *Am. chem. Soc. Monogr.* **182** (ed. Searle, C. E.) Ch. 4 (1984).

Upward fluxes of particulate organic matter in the deep North Pacific

K. L. Smith Jr*, P. M. Williams† & E. R. M. Druffel‡

Marine Biology Research Division, A-002 Scripps Institution of Oceanography, La Jolla, California 92093, USA

† Institute of Marine Resources, A-018 Scripps Institution of Oceanography, La Jolla, California 92093, USA

‡ Chemistry Department Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

The flux of particulate matter through the oceanic water column is a primary component in elemental cycling and is generally perceived as being in one direction: downward^{1,2}. The organic matter constituting these particles is produced through photosynthesis in surface waters and either sinks directly as phytoplankton and products^{3,4} or undergoes various trophic transformations through the water column. A large proportion of the particulate organic matter produced in surface waters is regenerated in the euphotic zone⁵⁻⁷. A fraction of this organic matter, however, leaves the surface waters and settles through the water column, generally decreasing in quantity and changing in quality with increasing distance from the surface⁸⁻¹¹. Although the net transport of organic matter must be downward to fuel the lower portions of the water column, there is also an upward component to transport. Positively buoyant particles, including lipid-rich eggs, larvae and, possibly, carcasses of deep-sea animals are examples of particles which undergo upward transport^{12,13}. A previous attempt to quantify the upward mass flux indicated rates of 1-4% of the downward mass flux¹⁴. Here we report the first evidence that there is a significant upward flux of particulate organic matter, up to 66.7% of the concurrently measured downward flux, at two stations in the deep North Pacific. Given this magnitude, the previously ignored upward flux of such organic matter must be considered in models of carbon and nitrogen cycling in the open ocean.

We concurrently measured the upward and downward flux of particulate organic matter (POM) at two stations in the North Pacific, each during two different seasons (Table 1). Station F was located in the relatively eutrophic eastern North Pacific on the western boundary of the California Current with an overall water of 4,400 m. Station CNP was in the oligotrophic central gyre north of the Hawaiian Islands with a water depth of 5,800 m¹⁵.

The downward flux of POM, that is, particulate organic carbon (POC) and particulate total nitrogen (PTN), was significantly higher ($P < 0.02$) at Station F than at Station CNP when compared at combined depths and seasons (Fig. 1). The downward fluxes were higher in the autumn at F and higher in spring at CNP. The carbon/nitrogen ratios of this sinking particulate matter were lower at F than at CNP. The distinguishable fraction of this sinking material was composed of small fecal pellets, diatom frustules, foraminiferan tests and a few pteropod shells.

The upward flux of POM (POC and PTN) was similar ($P > 0.05$) at both stations when compared at the same distance above the bottom and at combined depths and seasons. The upward fluxes were higher in spring at both stations. Carbon/nitrogen ratios of the positively buoyant POM were higher at F than at CNP where the autumn ratios were noticeably lower (Table 1). This buoyant material was largely composed of small (< 2 mm diameter) transparent spheres, some with orange pigment (possibly eggs), and less abundant gelatinous masses. Particularly noteworthy were 10 larvae of a bathypelagic euphausiid (*Thysanopoda* sp.)^{16,17} collected in the sediment trap 1,600 m above bottom in the spring at F (Table 1). These larvae (metanauplii and calyptopis I stages) do not have well developed appendages for swimming¹⁸ and appear dependent on their yolk stores for positive buoyancy. We decided to remove these animals from

our samples and they were not considered when calculating the upward flux of POM (see Table 1 legend).

Concurrently measured fluxes of POM (POC and PTN) were always higher in the downward than in the upward direction (Fig. 1). In the one concurrent data set of sufficient size for statistical comparison, at F, 1,600 m above bottom in spring, the POC and PTN fluxes were significantly higher ($P < 0.05$) in the downward direction. Combining both depth measurements in a single season revealed significantly higher downward fluxes ($P < 0.02$) of POC at F in spring and CNP in the autumn. Similarly, the downward fluxes of PTN were significantly higher ($P < 0.01$) than the upward fluxes at F in spring but not significantly different ($P > 0.05$) at CNP in the autumn. Carbon/nitrogen ratios of the upward flux of POM were consistently higher at F but always lower at CNP than the concurrently measured downward flux of POM.

Upward fluxes of POC constituted from 1.6% to 27% of the downward flux at F in the autumn and spring, respectively. Similarly, the upward flux of PTN ranged from 1.1% to 21.3% of the downward flux at the same station. In contrast, the upward flux of POM was more important at the oligotrophic central gyre station where the upward POC flux was 23.5–37.5% of the

downward flux, while the upward flux of PTN was 37.5–66.7% of the downward flux.

Our measured rates of downward fluxes of POM fall within the range of values measured previously at these same stations¹⁵ and are of similar magnitude to those reported previously for the eastern and central North Pacific^{19,20}. The only other measurements of upward flux of particulate matter in the open ocean were made on a 262-day deployment in an area ~750 km north-west of F at a similar depth (4,215 m)¹⁴. The mean mass flux was $0.7 \text{ mg m}^{-2} \text{ d}^{-1}$, and was rich in hydrocarbons, fatty acids, wax esters and ketones. They found the upward mass flux was from 1–4% of the concurrently measured downward mass flux. It is difficult to compare this study with ours, as the methods were not described and neither total organic carbon nor nitrogen were reported. The relative proportion of the upward to the downward flux is comparable, however, to those we measured in the autumn but substantially less than those found in the spring at F.

The higher upward fluxes we measured in the spring at both stations may be a result of seasonal reproductive efforts by deep-sea animals with planktotrophic development to take advantage of increased primary productivity in the overlying

Table 1 Flux of passively sinking (down) and rising (up) small particulate organic carbon (POC) and total nitrogen (PTN) measured at 600 and 1,600 m above bottom at F (4,400 m depth) and CNP (5,800 m depth)

Collection period	Moored sediment trap				POC flux (mg C m ⁻² d ⁻¹)				PTN flux (mg N m ⁻² d ⁻¹)				C/N	
Dates	Time (d)	Depth (m)	Altitude (m)	Current speed (cm s ⁻¹)	n	Down	n	Up	n	Down	n	Up	Down	Up
Station F (32°50' N, 124° W)														
Spring (1986)														
23 May-1 June	8.5	3,800	600	1.8 ± 1.0	1	1.03‡	1	0.08	2	0.17 ± 0.06	1	0.005	7.1	16.0
1 June-10 June	8.5	3,800	600		1	1.36‡	1	0.13			1	0.013		
23 May-1 June	8.5	2,800	1,600†	2.7 ± 1.7	2	0.42 ± 0.16	2	0.16 ± 0.06	2	0.05 ± 0.02	2	0.01 ± 0.00	8.4	16.0
1 June-10 Jan.	8.5	2,800	1,600†		2	1.06 ± 0.33	2	0.25 ± 0.08	2	0.12 ± 0.04	2	0.02 ± 0.01	8.8	12.5
Autumn (1986)														
18 Nov.-30 Nov.	11.0	3,800	600	1.6 ± 1.2	1	2.30‡	2	0.03 ± 0.00	1	0.28	2	0.003 ± 0.00	8.2	10.0
30 Nov.-13 Dec.	12.0	3,800	600		2	2.03 ± 0.19‡	2	0.04 1.33*	2	0.27 ± 0.05	2	0.004 0.16*	7.5	10.0 8.3
Station CNP (31° N, 159° W)														
Spring (1987)														
6 June-20 June	13.4	5,200	600	2.9 ± 2.0	1	0.87‡	2	0.17 ± 0.04	1	0.04	2	0.02 ± 0.01	21.8	8.5
20 June-3 July	12.1	5,200	600		2	0.49 ± 0.01‡	1	0.27	2	0.06 ± 0.001	1	0.03	8.2	9.0
Autumn (1985)														
8 Oct-22 Oct.	13.3	5,200	600	1.5 ± 1.2	1	0.20‡	1	0.13	1	0.02	1	0.02	10.0	6.5
22 Oct.-6 Nov.	14.0	5,200	600		1	0.28‡	1	0.04	1	0.03	1	0.01	9.3	4.0
11 Oct.-22 Oct.	11.0	4,200	1,600	1.8 ± 2.1	1	0.30	2	0.11 ± 0.08	1	0.03	2	0.02 ± 0.02	10.0	5.5
23 Oct.-6 Nov.	13.1	4,200	1,600		1	0.37	2	0.05 ± 0.00	1	0.05	2	0.005 ± 0.00	7.4	10.0

Upright and inverted conical sediment traps, each with a collection surface of 0.25 m^2 (ref. 27) and with sequencing collection cups¹⁵ were moored at 600 and 1,600 m above the bottom at each station for periods of 8.5 to 14 days. Collection cups at the bottom/top of each upright/inverted sediment trap were sequenced with collection only at the moored depth, freely flushing on ascent and descent. These collection cups work equally well in either an upright or inverted position because of an angled sidearm entry, equidistant between the 2 ends of the capped plexiglass cylinder. The collection cups were filled prior to deployment with water obtained from the depth of deployment and filtered through pre-combusted GF/C (Whatman glass fibre) filters. One trap of each tandem set was poisoned with mercuric chloride to retard degradation during the collection period. There was no significant difference ($P > 0.05$) between the POM flux measured in the poisoned and unpoisoned trap samples during the 12 collection periods, and hence they were considered duplicate samples. Samples of initial and final water from each collection cup were frozen for later determination of dissolved organic carbon and nitrogen to evaluate dissolution of particulate matter during incubation in the traps³¹. These samples were not analysed, however, because of the current controversy over an acceptable method³². The particulate samples from each collection cup were filtered through precombusted/pre-weighed GF/C filters and frozen for later analyses of POC and PTN in the laboratory. Before chemical analysis, whole animals ('swimmers') were removed from the filters. Filters from the upright traps were divided into quarters. One quarter of each filter was treated with HCl vapour to remove carbonates and analysed for organic carbon and total nitrogen using a Perkin-Elmer 240 CHN analyser. The remaining three-quarters of each filter was used in other analyses not reported here. Filters from the inverted traps were analysed whole as described above. Intra- and inter-station comparisons of POM fluxes were analysed with the Mann-Whitney U test with all significance levels chosen for two-tailed situations²⁸. An SIO model (no. 6) current meter was concurrently deployed with the sediment traps at each depth to estimate the trapping efficiency. The mean current speeds given in Table 1 suggest trapping efficiencies $> 80\%$ ²⁹.

* The (30 November–13 December) POM sample with a large gelatinous mass included in analysis.

† Each of these samples had euphausiid calyptopis I and metanauplius stages which were removed and are not included in these flux calculations.

‡ See ref. 30.

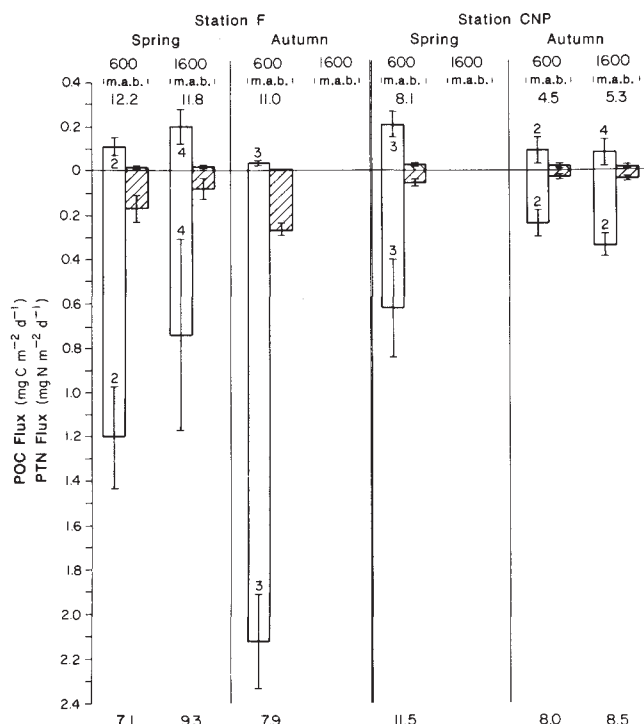


Fig. 1 Downward and upward fluxes of particulate organic carbon (POC) and particulate total nitrogen (PTN) measured at two stations, F and CNP in the North Pacific during spring and autumn at 600 and 1,600 m above the bottom (m.a.b.) (sampling was unsuccessful at 1,600 m.a.b. in autumn at F and spring at CNP). Upward fluxes are represented above the zero baseline and the downward fluxes below the baseline. The open bars represent the POC flux and the shaded bars the PTN flux. Each bar represents the mean value with the number of measurements $\pm$ one standard deviation (the number of measurements of POC flux associated with each bar also represents the concurrently measured PTN flux). The carbon/nitrogen ratios are given above the upward fluxes and below the downward fluxes.

waters. The large number of larvae from the bathypelagic euphausiid, *Thysanopoda* sp., collected in the inverted sediment trap at 1,600 m above the bottom in spring at F suggests a contribution of planktotrophic development to vertical fluxes. One mesopelagic species of *Thysanopoda* (*T. acutifrons*) spawns in late winter with larvae present during the spring in the North Atlantic²¹. The large contribution of wax esters to the upward flux suggests a crustacean source¹⁴. Planktotrophic development has been identified in a wide variety of deep-sea pelagic and benthic species^{22–26}. However, the only quantitative evaluation of this mode of development to the upward flux of POM has shown a minor contribution, based on studies of a dominant slope-dwelling species off California (manuscript in preparation).

Although our measurements have been made at abyssal depths, the importance of upward fluxes of POM need to be examined at shallower depths as well. The fluxes of POM have historically been considered in only one direction: downward. It is time that the fluxes of nutrients are re-evaluated in the context of biological processes associated with migrating animals and their products in upward and downward directions on a variety of temporal scales from diel to seasonal (for example, ontogenetic) migrations.

We thank R. Baldwin, J. Edelman, M. Laver, K. Robertson and R. Wilson for technical assistance during various stages of this research. The euphausiid larvae were identified by E. Brinton and M. Knight who enlightened us about these unique animals. F. Azam, R. Baldwin, A. Carlucci, R. Kaufmann and R. Eppley provided critical comments on the manuscript. This research was supported by grants from the NSF.

Received 21 November 1988; accepted 11 January 1989.

1. Dymond, J. & Lyle, M. *Limnol. Oceanogr.* **30**, 699–712 (1985).
2. Pace, M. L. *et al.* *Nature* **325**, 803–804 (1987).
3. Bienfang, P. K. *Mar. Biol.* **61**, 69–77 (1980).
4. Zeitschel, B. *Proc. R. Soc. Edin.* **88B**, 207–220 (1986).
5. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
6. Baker, E. T. *et al.* *Estuar. Coastal Shelf Sci.* **21**, 859–877 (1985).
7. Gardner, W. D. *et al.* *Mar. Geol.* **65**, 199–242 (1985).
8. Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **26**, 181–186 (1981).
9. Lee, C. & Cronin, C. J. *Mar. Res.* **40**, 227–251 (1982).
10. Lorenzen, C. J. *et al.* *Deep-Sea Res.* **30**, 639–643 (1983).
11. Matsueda, H. & Handa, N. *Mar. Chem.* **20**, 179–195 (1986).
12. Yayanos, A. A. & Nevenzel, J. C. *Naturwissenschaften* **65**, 255–256 (1978).
13. Robison, B. H. & Lancraft, T. M. *Naturwissenschaften* **71**, 322–323 (1984).
14. Simoneit, B. R. T. *et al.* *Naturwissenschaften* **73**, 322–325 (1986).
15. Smith, K. L. Jr. *Limnol. Oceanogr.* **32**, 201–220 (1987).
16. Brinton, E. J. *Crust. Biol.* **7**, 636–666 (1987).
17. Nemoto, T. *et al.* *Bull. Plankt. Soc. Jap.* **24**, 36–43 (1977).
18. Mauchline, J. & Fisher, L. R. *Adu. Mar. Biol.* **7**, 1–454 (1969).
19. Honjo, S. *J. Mar. Res.* **38**, 53–97 (1980).
20. Tsunogai, S. & Noriki, S. *Deep-Sea Res.* **34**, 755–767 (1987).
21. Mauchline, J. *Mar. Biol.* **90**, 19–26 (1985).
22. Merrett, R. *Deep-Sea Res.* **25**, 147–160 (1978).
23. Marshall, N. B. *Developments in Deep-Sea Biology* (Blandford, Poole, 1979).
24. Tyler, P. A. & Gage, J. D. *Deep-Sea Res.* **31**, 387–402 (1984).
25. Colman, J. G. *et al.* *J. Mar. Biol. Ass. U.K.* **66**, 951–965 (1986).
26. Killingley, J. S. & Rex, M. A. *Deep-Sea Res.* **32**, 809–818 (1985).
27. Bruland, K. W. *et al.* *Earth planet. Sci. Lett.* **53**, 400–408 (1981).
28. Conover, W. J. *Practical Nonparametric Statistics* (Wiley, New York, 1980).
29. Gardner, W. D. *J. Mar. Res.* **38**, 17–39 (1980).
30. Smith Jr, K. L. *Deep-Sea Res.* **36** (in the press).
31. Knauer, G. A. *et al.* *J. Mar. Res.* **42**, 445–462 (1984).
32. Williams, P. M. & Druffel, E. R. M. *Oceanography* **1**, 14–17 (1988).

Direct observation of a section through slow-spreading oceanic crust

J.-M. Auzende*, D. Bideau*, E. Bonatti†‡, M. Cannat§, J. Honnorez||, Y. Lagabrielle§¶, J. Malavieille#, V. Mamaloukas-Frangoulis§ & C. Mevel**

* IFREMER CB, BP70 29263 Plouzané, France

† Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA

‡ Dipartimento Scienze della Terra, Pisa, Italy

§ GIS Océanologie et Géodynamique, Avenue Le Gorgeu, 29283 Brest Cedex, France

|| Institut de Géologie, Rue Blessig, 67000 Strasbourg, France

USTL, Laboratoire de Tectonique, Place E. Bataillon, 34060 Montpellier Cedex, France

** UPMC, Laboratoire de Petrologie—UA040736, Place Jussieu, 75252 Paris Cedex, France

Understanding the nature and composition of the oceanic crust has been a longstanding goal of Earth scientists. Seismic refraction experiments^{1–3} suggest a simple layered crust made of eruptive basalts underlain by a thick layer of doleritic and gabbroic intrusives and a peridotitic upper mantle. Other evidence comes from ophiolite complexes on land⁴, although generalizations based on ophiolites are uncertain because they may be dismembered and altered during emplacement, and it is not known whether they represent sections of 'mature' oceanic crust, or crust from very small 'aborted' oceans⁵, anomalous ocean structures⁶ or marginal basins. The walls of fracture-zone valleys expose thick sections of oceanic lithosphere which are accessible to *in situ* observations and sampling^{7,8}, but this approach has been criticized because the pattern of faulting in fracture zones may disrupt the original stratigraphy of the crust⁹, and because the crust near fracture zones is anomalously thin^{3,10,11}. Here we report the direct observation and sampling of a section of crust and upper mantle exposed at the Vema fracture zone in the Atlantic, using the French submersible *Nautil*.

The Vema transform fault¹² offsets left laterally the axis of the Mid Atlantic Ridge (MAR) by about 320 km at 11°N

¶ Permanent address: URA Géodynamique, Parc Valrose, 06000 Nice, France.

(Fig. 1). It is one of the largest seismically active equatorial Atlantic transforms. Given the low spreading rate of the lithosphere in the vicinity of the transform ($\sim 1.2 \text{ cm yr}^{-1}$), the slip rate of the North American and African plates across the Vema transform is about 2.4 cm yr^{-1} (refs 7, 13–15).

The Vema fracture zone is characterized by a deep (5,100 m) east-west valley filled with sediments which reach up to 1.5 km in thickness. It is bounded by steep E-W walls which rise 2–4 km above the valley floor^{13,14}. The southern wall marks the flank of an E-W (transverse) ridge running parallel to the valley. This transverse ridge becomes very elevated (up to $\sim 600 \text{ m}$ depth) towards the western section and constitutes a major topographic anomaly relative to the predicted depth/age relationship from the lithospheric thermal-subsidence curve^{14,16}. It has been subjected to strong vertical movements since at least Miocene time and it is probably an uplifted but otherwise unchanged part of crust and upper mantle^{16–18}.

Rocks dredged from the fracture-zone valley walls include a variety of mantle-derived peridotites, gabbros, dolerites and basalts^{7,8,19,20}, that is, samples of all the major units that are inferred, by analogy with ophiolites, to constitute the oceanic crust. Present dredging data do not, however, permit an accurate assessment of the stratigraphic relationships among these different crustal units.

Two north-south geological transects were carried out with the submersible *Nautille* of IFREMER on the southern wall of the Vema transform (transects AB and CD, Figs 1 and 2). The objective of this survey was to evaluate, by direct detailed observation and sampling, whether or not the different units exposed on the southern wall amount to a complete section of the oceanic crust. These parallel transects are located about 5 km apart around $42^\circ 42' \text{ W}$, in an area where serpentinized peridotites and amphibolites have previously been dredged²⁰. Five dives were carried out, covering the entire height of the southern wall, from the fracture-zone valley floor at 5,150 m depth to the crest of the transverse ridge between 2,100 and 2,280 m depth (Fig. 2). Both profiles were initiated in the flat sediment at the foot of the wall. East-west alignments of live clam colonies, that is, parallel to the scarp, were observed in the sediments at two levels, suggesting the discharge of hydrothermal solutions derived from sub-seafloor-circulating sea water, possibly along E-W faults.

The following units were observed and sampled, starting from the base of the section.

Mantle-derived ultramafics. Serpentinized peridotites were the only rocks outcropping from the sediment cover below 4,200 m below sea level (b.s.l.), that is, in the lower 1 km or so of both sections (Figs 2, 3), except for fault-bounded isolated outcrops of amphibolites at 4,800 m b.s.l. and of gabbro at 4,400 m, observed during dive 8. The outcrops are found in part tectonically disrupted and brecciated; they are affected by N-S diaclasses and by small E-W to $N60^\circ\text{--}90^\circ$ strike-slip faults, locally associated with peridotitic mylonites. The ultramafics were sampled with the submersible's mechanical arm at 14 sites. Most of the rocks have porphyroclastic texture; some are mylonitic. They are highly serpentinized but contain relicts of olivine, orthopyroxene, and occasionally clinopyroxene and spinel, suggesting that they last equilibrated under conditions of pressure and temperature appropriate to spinel peridotite facies, that is under a pressure of at least 8–9 kbar (corresponding to a depth of about 30 km). The mylonitic peridotite contains, in addition to olivine, orthopyroxene and clinopyroxene, minor quantities of plagioclase, implying that recrystallization occurred at lower pressure ($<9 \text{ kbar}$). Therefore, texturally and mineralogically these peridotites are similar to mantle-derived peridotites sampled from the Vema fracture zone in the past^{7,19}.

Gabbros. At depths from about 4,200 to 3,700 m b.s.l. the two profiles encountered a gabbroic unit. Towards the base of this unit the exposures are aligned N-S, and towards the top the outcrops are more massive. Samples collected from 11 sites

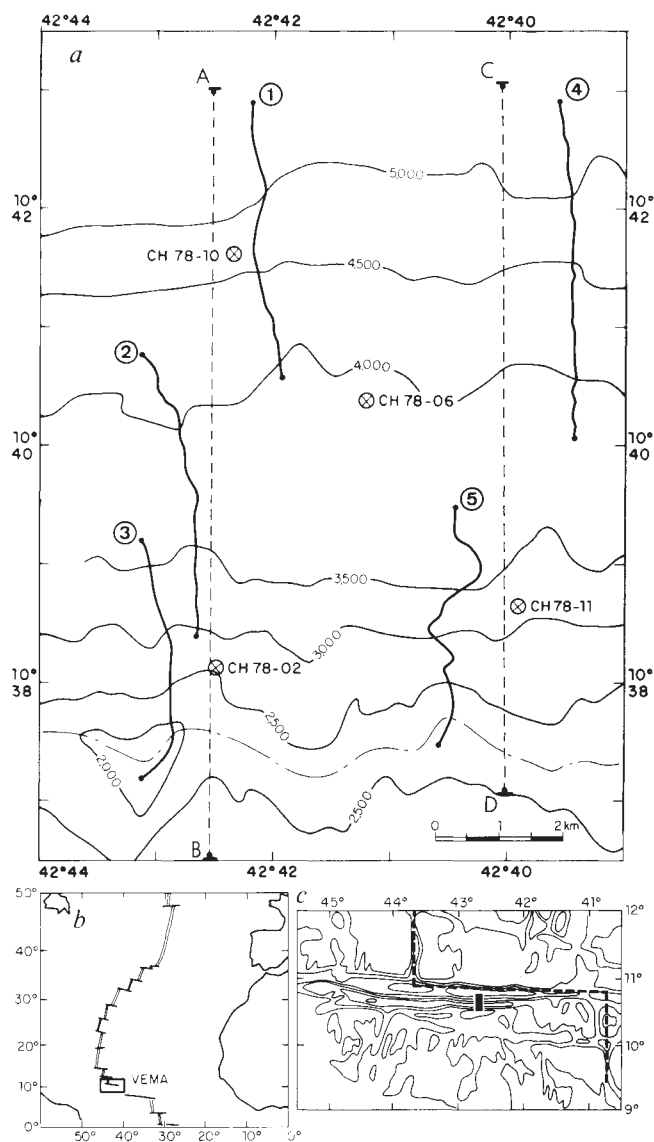


Fig. 1 Location maps of *a*, the Vemaute dives (CH78-02 denotes the Jean Charcot dredgings; A-B and C-D show the synthetic profiles); *b*, the Vema fracture zone in the North Atlantic framework; and *c*, schematic bathymetry of the Vema fracture zone. The surveyed area is shown as a black rectangle, and dashed lines denote the American-African plate boundary.

within this unit indicate that the exposed rocks are predominantly Fe- and Ti-rich gabbros.

Dyke complex. The top of the gabbro unit is cut by doleritic dykes which are overlain on the scarp by an approximately 50-m-thick level of semi-consolidated, sub-horizontal sediment. Above 3,500 m and up to about 2,400 m a dyke complex outcrops almost continuously for a thickness of over 1 km, being interrupted only by a few sub-horizontal or chaotic lava flows. In the lower part of this section packets of dykes are partly disrupted by mass wasting; in the upper part we observed thick outcrops of relatively undisturbed dyke complex, exhibiting nearly complete dyke exposure and characteristic 'dyke-in-dyke' structures²¹ (Fig. 3). Samples collected at 12 sites within this unit consist of basaltic dolerites with various crystal sizes, displaying various stages of alteration. The dykes (Fig. 3) are vertical or dip slightly to the east; they are orientated $N-S \pm 30^\circ$, that is, parallel to the strike of the present MAR axis and perpendicular to the transform. This orientation is consistent with the

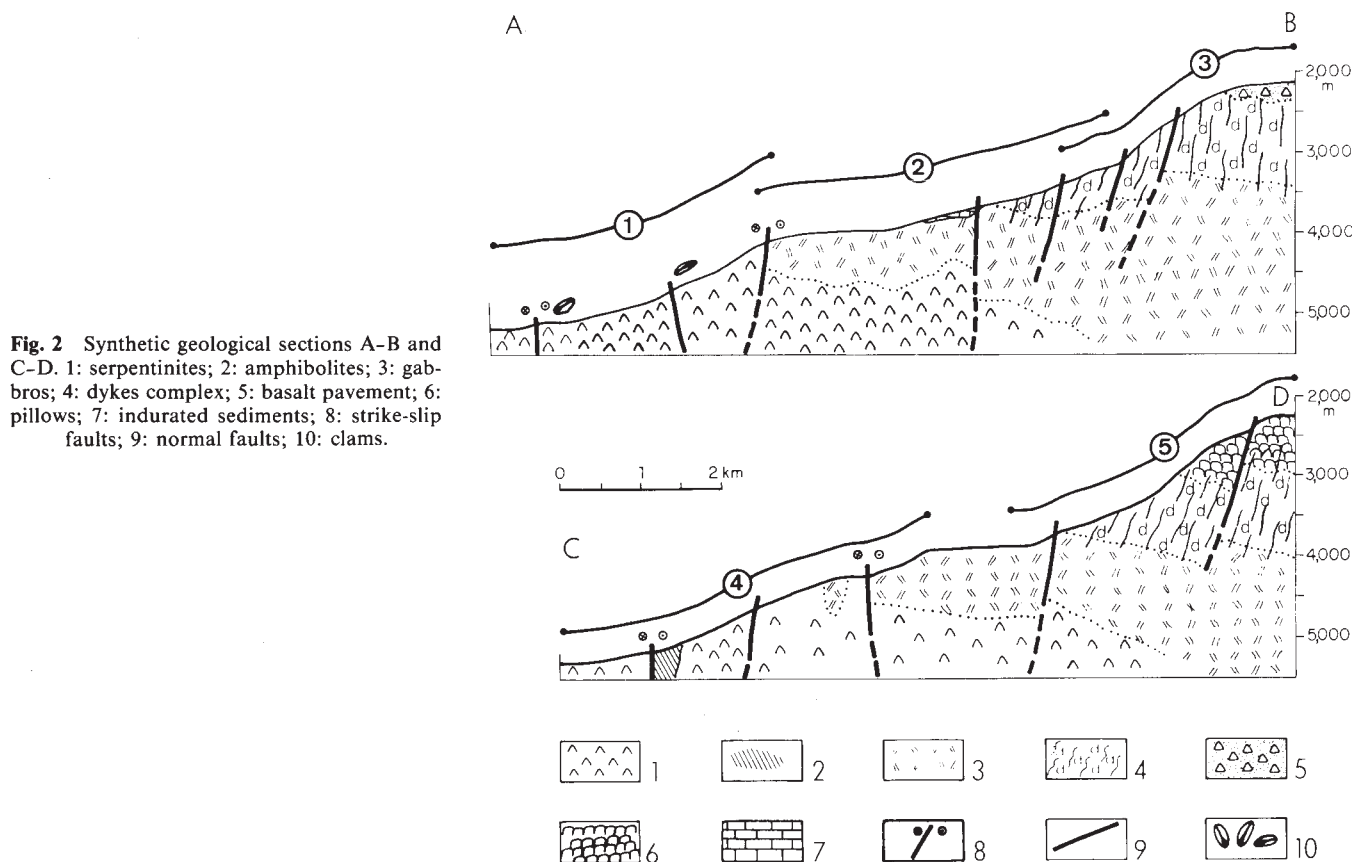


Fig. 2 Synthetic geological sections A-B and C-D. 1: serpentinites; 2: amphibolites; 3: gabbros; 4: dykes complex; 5: basalt pavement; 6: pillows; 7: indurated sediments; 8: strike-slip faults; 9: normal faults; 10: clams.

hypothesis that the dyke complex was originally emplaced along the active, axial zone of the MAR. This is, to our knowledge, the first time that a true dyke complex has been observed in present-day oceans.

Basaltic upper crust. A few small sub-horizontal basaltic flows are scattered within the dyke complex. At about 2,400 m b.s.l. we observed in both profiles a transition to a dominantly extrusive basalt unit which includes largely disrupted and brecciated pillow- and pavement-type lavas. This unit covers the crest of the transverse ridge along the two profiles and is up to 800 m thick. Samples collected from this sequence at five sites consist of weathered basalt and fine-grained dolerite.

In summary, the sequence exposed on the southern wall of the Vema fracture-zone valley at around 42°42' W consists of (from bottom to top) (1) upper-mantle-derived peridotites (exposed for a thickness of about 1 km); (2) a lower-crust gabbroic unit (thickness of about 500 m); (3) a mid-crust dyke complex (thickness of 700–1,100 m); and (4) an upper-crust eruptive basalt unit (thickness of up to 800 m). This sequence is quite similar to those of present models of the ocean crust, as discussed below.

These profiles may be summarized in terms of two geological sections (Fig. 2), which take into account the set of E-W, presumably normal, faults observed along the profiles, as well as the apparent strike-slip tectonic structure. If the Vema crustal section was originally created at the eastern intersection of the transform and the MAR axis (within the present ridge-axis/transform geometry), its age is ~16–17 Myr, assuming a spreading rate of 1.2 cm yr^{-1} (refs 7, 14). Alternatively, if transform migration and reorientation occurred about 10 Myr ago, as has been suggested²², the Vema section could have formed at the ridge axis but away from its intersection with the transform. K/Ar ages of ~10 Myr for amphibolites dredged close to our section²⁰ are consistent with the first possibility, unless

reheating during the transform migration of the second suggestion has reset the radiometric clock.

The observed sequence is remarkably similar to those inferred from seismic-refraction experiments, deep-sea drilling and observation of ophiolites. DSDP hole 504B drilled a 1,287-m-thick section into the Nazca plate crust, including 571 m of basalt flows, 209 m of a lava-flow/dyke transition zone and over 500 m of dolerite²³.

The thickness on the Vema section (Fig. 2) of the lower-crustal gabbroic layer (~500 m) is less than that in the reference section and in most models of the oceanic crust^{1,2,24}. One possible explanation might be that the lower part of the gabbroic layer has been covered by a diapiric rise of slivers of serpentinized peridotite along major transform-parallel faults, reducing its total apparent thickness. This possibility is supported by the



Fig. 3 Photograph of the dykes complex. Dive no. 3; 2,775 m depth. The area shown in the photograph is approximately 5 × 5 m.

fact that we find in the Vema section Fe- and Ti-rich gabbros, which presumably represent the upper part of the gabbroic sequence, but not Mg gabbros or ultramafic cumulates, which should be abundant in the lower part of the gabbro layer.

The extrusive upper-crustal basalt layer is somewhat thinner in the Vema section (~800 m) than is indicated by most models of the oceanic crust, in which typically it reaches ~1 km. Mass wasting during transport along the edge of the fracture-zone valley might have reduced the thickness of this layer. The dyke-complex thickness of about 700–1,100 m is within the limits of most models of oceanic crust.

Above the peridotites, the thickness of the Vema crustal section is roughly 2.2 km. To this we should add the basal part of the lower-crustal gabbroic layer, which is inferred to be missing, and the upper part of the basalt layer, and we should subtract the repetitions of the series that arise from normal faulting. Although these factors are difficult to assess, we estimate that the original crustal thickness of the Vema section was somewhat greater than the 2.2 km exposed, and is compatible with the seismic structure^{25,26} of the fracture.

In conclusion, two profiles on the southern wall of the Vema fracture-zone valley (the north side of the transverse ridge) demonstrate that transform faults can expose relatively thick sections of oceanic crust, contrary to previous suggestions⁹.

We thank Captain Chatelet and the crew of RV *Nadir* as well as P. H. Nargeolet and the *Nautilie* Team. The Vema-naute cruise was founded by IFREMER. We also thank H. D. Needham, K.

Kastens and K. Macdonald for providing useful documents, and B. Collette, J. Karson and K. Loudon for comments on the manuscript.

Received 17 October 1988; accepted 13 January 1989.

1. Raitt, R. W. in *The Sea* Vol. 3 (ed. Hill, M. N.) 85–102 (Wiley, New York, 1963).
2. White, R. S. in *Ophiolites and oceanic lithosphere* (eds Gass, I. G., Lippard, S. J. & Shelton, A. W.) Geol. Soc. spec. Publ. no. 13, 101–111 (1983).
3. Detrick, R. S. & G. M. Purdy, *J. geophys. Res.* **85**, 3759–3777 (1980).
4. Coleman, R. *Ophiolites: ancient oceanic lithosphere?* (Springer, Berlin, 1977).
5. Lagabrielle, Y. thesis, Univ. Brest (1987).
6. Auzende, J. M., Polino, R., Lagabrielle, Y. & Olivet, J. L., *C. R. Acad. Sci. Paris* **296**, 1527–1532 (1983).
7. Bonatti, E. & Honnorez, J. *J. geophys. Res.* **81**, 4104–4116 (1976).
8. Bonatti, E., Honnorez, J. & Ferrara, G. *Phil. Trans. R. Soc. Lond.* **268**, 385–402 (1971).
9. Francheteau, J., Choukroune, P., Hekinian, R., Le Pichon, X. & Needham, H. D. *Can. J. Earth Sci.* **13**, 1223–1235 (1975).
10. Fox, P. J. & Gallo, D. G. *Tectonophysics* **104**, 205–242 (1984).
11. White, R. S., Detrick, R. S., Sihna, M. C. & Cormier, M. H. *Geophys. J. R. astr. Soc. Lond.* **79**, 779–798 (1981).
12. Heezen, B. C., Gerard, R. D. & Tharp, M. *J. geophys. Res.* **69**, 733–739 (1964).
13. Van Andel, T. H., Corliss, J. B. & Bowen, V. T. *J. mar. Res.* **25**, 343–351 (1967).
14. Van Andel, T. H., von Herzen, R. D. & Phillips, J. D. *Mar. geophys. Res.* **1**, 261–283 (1971).
15. Roest, W. R. & Collette, B. J. *J. geol. Soc. Lond.* **143**, 833–843 (1986).
16. Bonatti, E., Sartori, R. & Boersma, A. *Tectonophysics* **91**, 213–232 (1983).
17. Bonatti, E. & Honnorez, J. *Science* **174**, 1329–1331 (1971).
18. Honnorez, J. *et al. Earth planet. Sci. Lett.* **26**, 8–12 (1975).
19. Melson, W. G. & Thompson, G. *Phil. Trans. Roy. Soc. Lond.* **268**, 423–441 (1971).
20. Honnorez, J., Mevel, C. & Montigny, R. *J. geophys. Res.* **89**, 11379–11400 (1984).
21. Bonatti, E. & Crane, K. *Nature* **300**, 343–345 (1982).
22. Anderson, R. N. *et al. Nature* **300**, 589–594 (1982).
23. Karson, J. A. & Elthon, D. *Geology* **15**, 127–131 (1987).
24. Kidd, R. & Cann, J. R. *Earth planet. Sci. Lett.* **24**, 151–155 (1974).
25. Loudon, K. E., Potts, R. S., Potts, C. G. & Forsyth, D. W. *J. geol. Soc. Lond.* **143**, 795–805 (1986).
26. Potts, C. G., White, R. S. & Loudon, K. E. *Geophys. J. R. astr. Soc. Lond.* **86**, 491–513 (1986).

Evidence for a lost ocean in Variscan terranes of the western Massif Central, France

G. Dubuisson, J.-C. C. Mercier, J. Girardeau & J.-Y. Frison

Laboratoire Pétrologie Physique, Institut de Physique du Globe de Paris & Université de Paris 7, 75252 Paris Cedex 05, France

Plate tectonics is now commonly invoked to explain the Variscan orogeny^{1–3}, but only the presence of ophiolites can provide real evidence for an early Variscan ocean. None has been described so far in south-western Europe, with the possible exception of a few scattered lenses⁴ within the low-grade Moeche unit in Cabo-Ortega, Spain, and the Chamrousse ophiolite^{5,6}, French Alps, which displays no clear relationships to documented Variscan terranes. To fill this gap, eclogites with mid-ocean-ridge-basalt (MORB) compositions have been tentatively ascribed to an oceanic origin⁷, and various serpentinites and amphibolites have been inferred to derive from ophiolites^{1–3}, but many geologists are still reluctant to admit that the Variscan orogeny resulted from the closure of a true ocean. Here we describe the remains of an ophiolite found in high-grade crystalline rocks in the European Variscan fold belt, in Limousin, western French Massif Central. This ophiolite comprises diopside-bearing harzburgites, depleted harzburgites, dunites, wehrlites and MORB-type gabbros and has structural, petrological and geochemical characteristics typical of oceanic lithosphere, thereby providing evidence for a lower Palaeozoic ocean.

The Limousin comprises several crystalline nappes (Fig. 1) with different lithologies and metamorphic histories^{8–10}. One of these, the Intermediate Allochthon, contains ophiolite bodies (Fig. 2) first interpreted as layered intrusions^{11,12}, but does not bear any relic of former eclogites, which are common in the overlying Upper Allochthon⁶. According to the evidence from both mineral assemblages and zircon morphology¹³, meta-

morphism in the Intermediate Allochthon was limited to upper-amphibolite to granulite conditions.

The ophiolite comprises a subcontinuous belt, about 60 km long, of lenses (identified by letters in Fig. 1). Some of these (B, E and F, Fig. 3) show a nearly complete ophiolite sequence of harzburgites, dunites, wehrlites and gabbros. All of the peridotites contain secondary pargasitic hornblende, an amphibole which, according to textural evidence, does not derive from pyroxenes, and which is found even in the pyroxene-free dunites. Hence, it has to be related to a late penetrative modal metasomatism, which occurs with virtually no re-equilibration of the primary phases. This amphibole will therefore be neglected in the following, as will be all low-grade metamorphic products.

The structural thickness of the diopside-bearing harzburgites may locally reach 2 km (at lens A for example). The primary mineral assemblage (olivine, orthopyroxene, clinopyroxene and spinel) is generally well preserved, incorporating coarse orthopyroxenes similar to those found in the lower harzburgites of typical ophiolites. The primary foliation is homogeneous and conformable; pyroxenite layering is thin, when present. The textures range from fine-porphyroclastic to mylonitic. In basal mylonites, the orthopyroxenes are strongly elongated, with a length:width ratio of up to 20:1 (A). The fabrics for the olivine neoblasts result from activation of the (010)[100] slip system (Fig. 5). Hence, because of the high stresses (500 MPa) inferred from the grain-size piezometer¹⁴ for rotation recrystallization⁴, this deformation was produced at high temperature¹⁵ (>1,000 °C). These mantle temperatures (possibly corresponding to an early stage of obduction) are incompatible with a crustal environment⁹. The primary phases have homogenous compositions typical of ophiolitic harzburgites, with an aluminous spinel ($Al\# = Al/(Al+Cr) = 0.91$), a magnesian olivine ($Mg\# = Mg/(Mg+Fe) = 0.898$), a high-temperature aluminous and calcic enstatite and a residual Cr-rich diopside. The compositions of the primary clinopyroxenes indicate that equilibrium with nearby orthopyroxenes had been reached when these minerals last coexisted with a liquid, under estimated conditions of 1,050–1,200 °C and 0.8–1.2 GPa (refs 16, 17).

The depleted harzburgites are virtually undeformed, but are often totally retrogressed. Their original texture is well preserved

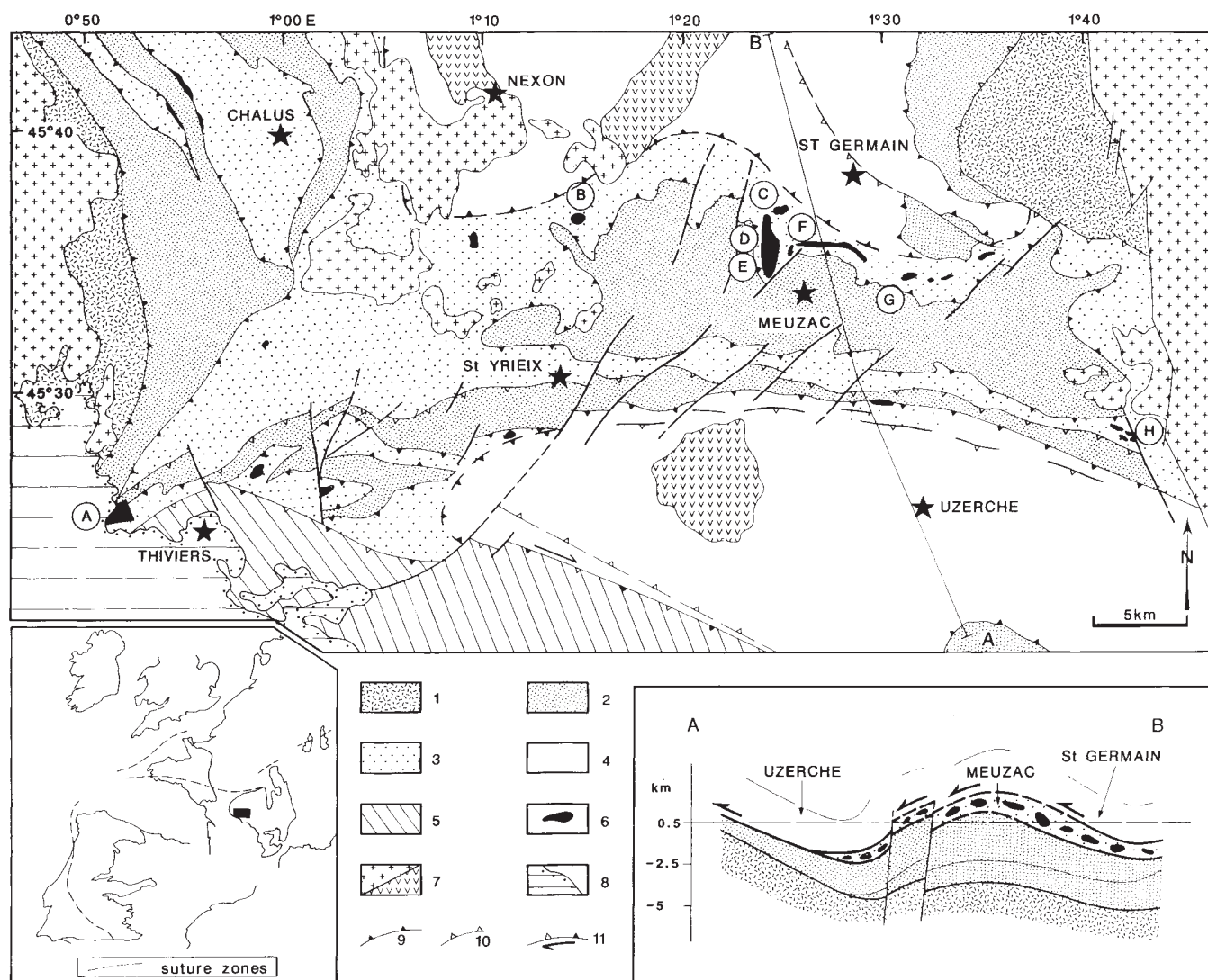


Fig. 1 Ophiolite location, geological sketch map and schematic cross-section of Limousin, western French Massif Central. The key is as follows: (1) relative autochthon; (2) lower/basal allochthon; (3) intermediate allochthon; (4) upper allochthon; (5) Thiviers-Payzac and Génis units; (6) ophiolite; (7) granite/diorite; (8) Aquitan sedimentary basin; (9) early thrust; (10) late ductile thrust; (11) late thrust. A: La Queue d'Ane; B: La Roche l'Abeille; C: Flaquecorne; D: La Flotte; E: Le Cluzeau; F: La Plagne; G: La Porcherie; H: Le Lonzac. Insets show geographical location, with suture zones indicated, and A-B cross-section.

and incorporates smaller orthopyroxenes with irregular lobed shapes (B, E), similar to those in upper harzburgites from ophiolites. Olivine has almost the same $Mg\#$ as that of the diopside-bearing harzburgites.

Massive dunites are highly serpentinized and are always tectonically shortened. They crop out either as decametric patches within the depleted harzburgites or as a distinct overlying unit (100 m thick at H). The dunites are residual according to their olivine and spinel compositions ($Mg\#_{ol} = 0.9095$; $Al\#_{sp} = 0.72$) and the occasional observation of a subfresh exsolution bearing orthopyroxene. Locally (B), they show ghosts of euhedral chromite typical of podiform-type deposits.

Wehrlites usually overlie the massive dunites, with an irregular primary contact, when visible. They show large variations in thickness at the 100-m scale, forming discontinuous patches no more than 400 m thick which would represent discrete magmatic zones, thus defining an irregular palaeo-Moho (D, E). The wehrlites usually have isotropic textures and are generally very fresh. Plagioclase and clinopyroxene show large variations in both size and distribution, sometimes to the point of defining patches of dunite and local magmatic stratification. Olivine

(locally 50% modal) is generally equant (2–3 mm) with large kink bands and is richer in Fe ($Mg\# = 0.857$) than are the harzburgites. Clinopyroxene (up to 60% modal) occurs as large poikilitic interstitial-like crystals with wide exsolution lamellae. It is less sodic in wehrlites than in harzburgites and is richer in chromium, as is spinel; this is ascribed to the presence of a plagioclase buffer (less than 15% modal) which occurs interstitially. Some primary amphibole has also been chemically identified.

The total lack of orthopyroxene in the wehrlites, which rest upon the dunites (in contrast to the orthopyroxene- and plagioclase-bearing wehrlites, which directly overlie harzburgites near Chaluz (Fig. 1), suggests that these wehrlites result from impregnation of the residual dunites by mafic liquids. This is consistent with the large volume of the wehrlites, which is difficult to explain on the basis of fractional crystallization from the standpoint of simple mass-balance. Therefore, impregnation would result in crystallization of clinopyroxene and plagioclase, and simultaneous chemical re-equilibration of both spinel and olivine (iron enrichment).

The gabbros overlie the wehrlites as thin superimposed tec-

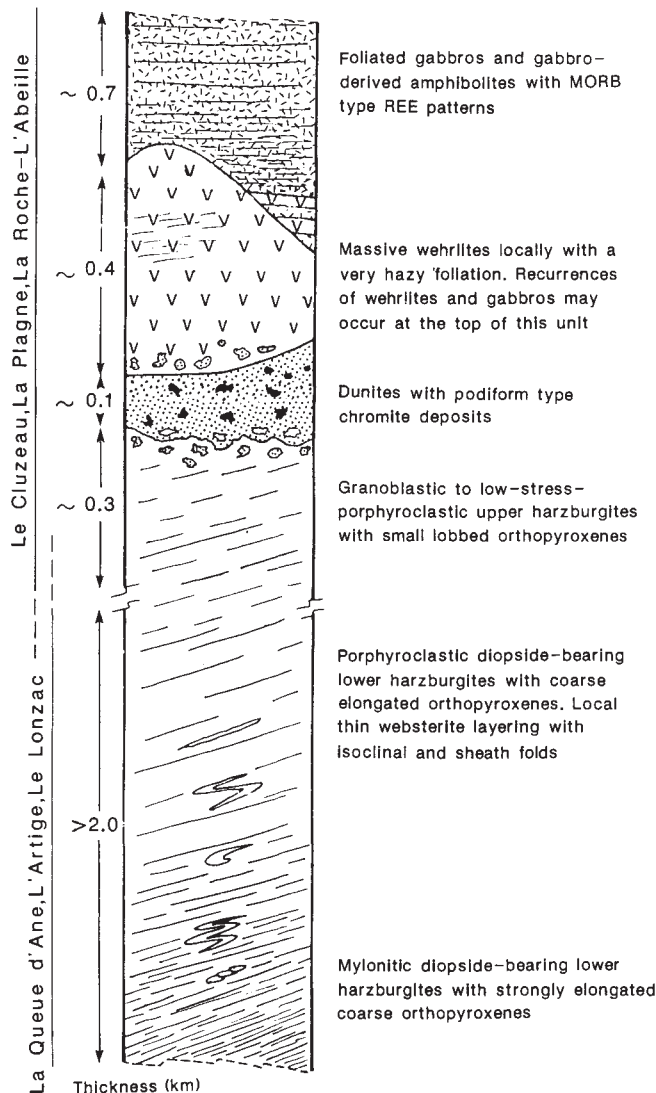


Fig. 2 Synthetic section of the Limousin ophiolite.

tonic flakes with a thickness locally reaching 700–800 m (D, E). They have been sheared into 'flaser' gabbros and foliated amphibolites. Most of the gabbros, including all those resting directly on the peridotites, are slightly greenish and at present contain a calcic plagioclase (An_{50}) and a magnesian hornblende (~50% model). Their rare-earth-element (REE) patterns (Fig. 4) show a depletion in light REE relative to chondrites Ce/Sm ratio <1), which is also the case in oceanic MORB tholeiites and is typical of the upper isotropic gabbros from ophiolites. The europium enrichment indicates a cumulative origin for the plagioclase. Locally (D), these greenish oceanic gabbros are tectonically overlain by greyish light-REE-enriched tholeiitic gabbros. Such associations have been found in geodynamic environments such as incipient¹⁸ and mature¹⁹ oceanic ridges, but these gabbros could alternatively be a tectonic flake of the Upper Allochthon, which contains eclogitized light-REE-enriched mafic rocks²⁰.

Thus, the Limousin area exhibits remnants of an ophiolite sequence that comprises diopside-bearing harzburgites, depleted harzburgites, dunites, wehrlites and massive gabbros crystallized from MORB magmas. The only radiometric data for this ophiolite consists of an Archaean 2.89-Gyr U–Pb age obtained for magmatic zircons from the gabbros^{13,22}. This age, which is obviously irrelevant to the Variscan orogeny, is still not understood.

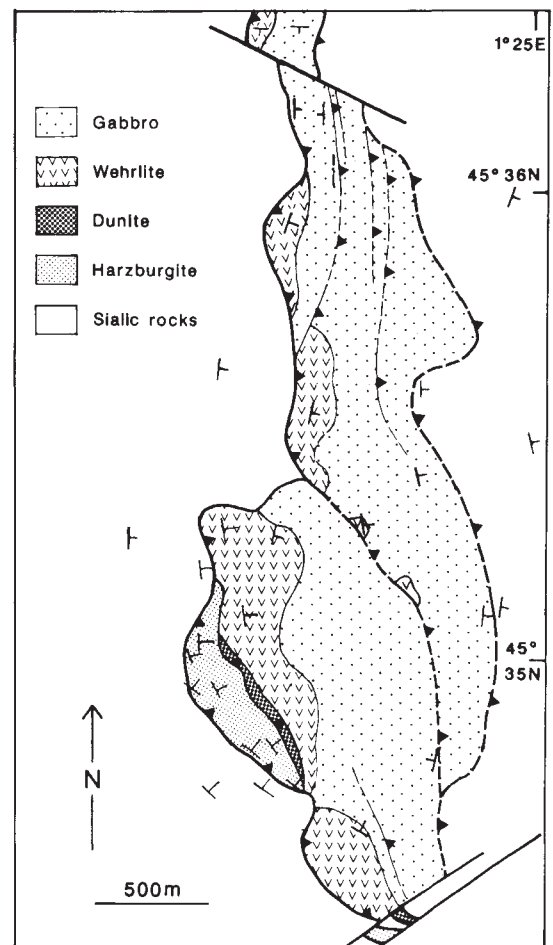


Fig. 3 Geological map of the Le Cluzeau–La Flotte massif.

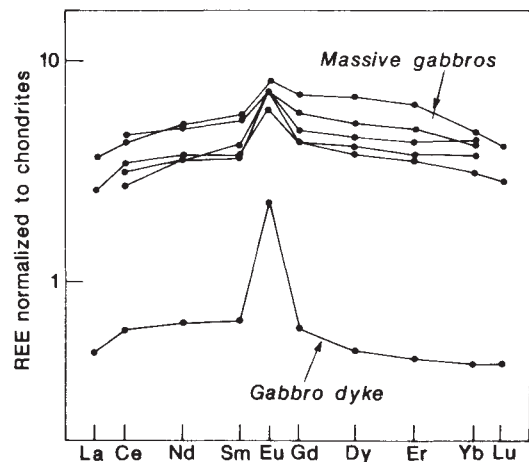
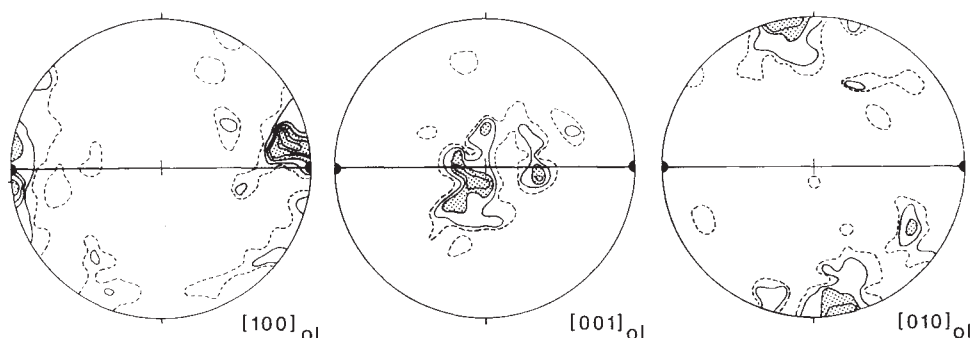


Fig. 4 REE patterns of tholeiitic gabbros from Limousin. Data obtained from isotope dilution by A. Prinzhofer.

If the ultramafic residual part of the ophiolite is well preserved, the crustal part has been strongly tectonized and possibly greatly reduced in thickness. Indeed, no layered cumulative gabbros, dolerites or basalts have so far been recognized in the studied area. It is not clear, however, whether evidence for these terms has been blurred by late tectonics, or whether lack of these terms is a primary feature, as it is in ophiolites formed at slow-spreading ridges^{20,23,24} and/or at the early stage of ocean formation^{24–26}.

Fig. 5 Olivine preferred orientation. Sample is mylonite from La Queue d'Ane, density contours based on 123 data; intervals at 1, 2, 4, 6 and 8% per 1% area of random orientation.



The Limousin ophiolite probably reached its present position in the early Carboniferous^{2,3}. It has been squeezed between an Ordovician orthogneiss (lower/basal allochthon, 470–495 Myr (refs 13, 27) and the Upper Allochthon, which contains eclogites similar to those of Silurian age in Haut-Allier and Vendée^{28,29}. Hence this ophiolite may be regarded as a relict of a lower Palaeozoic ocean, which is consistent with the 496-Myr age obtained for the probably subcontemporaneous Chamrousse crustal ophiolite⁶ and with the 500-Myr age for the early distension that ultimately yielded this ocean³⁰.

This work was supported by the French CNRS-BRGM (G.P.F.). We are grateful to A. Prinzhofer and E. Lesur for their technical help.

Received 24 June 1988; accepted 6 January 1989.

- Burg, J.-P. & Matte, P. Z. *Dt. Geol. Ges.* **129**, 429–460 (1978).
- Matte, P. *Tectonophysics* **126**, 329–374 (1986).
- Burg, J.-P., Bale, P., Brun, J.-P. & Girardeau, J. *Geodinamica Acta* **1**, 71–87 (1987).
- Arenas, R. *et al. Hercinica* **7** (in the press).
- Bodinier, J. L., Dupuy, C., Dostal, J. & Carme, F. *Contr. Miner. Petrol.* **78**, 379–388 (1981).

- Ménot, R. P., Peucat, J. J., Scarenzi, D. & Piboule, *Earth planet. Sci. Lett.* **88**, 82–92 (1988).
- Montigny, R. & Allègre, C. J. *C.R. Acad. Sci. Paris*, **279D**, 543–545 (1974).
- Floc'h, J.-P. thesis, Univ. Limoges (1983).
- Girardeau, J., Dubuisson, G. & Mercier, J.-C. *C. Bull. Soc. geol. France* **8** (II), 849–860 (1986).
- Dubuisson, G., Hirn, A., Girardeau, J., Mercier, J.-C. & Veinante, J.-L. *Tectonophysics* **147**, 19–31 (1987).
- Chenevoy, M., Forestier, F. H. & Gay, M. *C.R. Carte géol. France* (Bureau de Recherche Géologique et Minière, Orléans, France, 1962).
- Santalier, D. thesis, Univ. Orléans (1981).
- Lafon, J.-M. thesis, Univ. Montpellier (1986).
- Mercier, J.-C. *C. J. geophys. Res.* **85**, 6293–6303 (1980).
- Mercier, J.-C. in *Preferred orientation of minerals and rocks: an introduction to modern textural analysis* (ed. Wenk, H. R.) 407–430 (Academic, New York, 1985).
- Bertrand, P. & Mercier, J.-C. *Earth planet. Sci. Lett.* **76**, 109–122 (1985).
- Mercier, J.-C. C., Benoit, V. & Girardeau, J. *Contr. Miner. Petrol.* **85**, 391–403 (1984).
- Kornprobst, J., Vidal, Ph. & Malod, J. *C.R. Acad. Sci. Paris* **306**, Sér. II, 1359–1364 (1988).
- Prinzhofer, A. thesis, IGP & Univ. Paris (1987).
- Cabanis, B. *et al. Bull. Soc. geol. France* **7** (XXV), 563–574 (1983).
- Nicolas, A. *et al. Nature* **294**, 414–417 (1981).
- Lafon, J.-M. & Lancelot, J. R. *Terra Cognita* **6**, 156 (1986).
- Girardeau, J., Mercier, J.-C. & Zao Yougong *Tectonophysics* **119**, 407–433 (1986).
- Nicolas, A. *Rev. Geophys.* **24**(4), 875–895 (1986).
- Boudier, F. & Nicolas, A. *Earth planet. Sci. Lett.* **76**, 84–92 (1985).
- Lemoine, M., Tricart, P. & Boillot, G. *Geology* **15**, 622–625 (1987).
- Bernard-Griffiths, J. A. *Fac. Sci. Univ. Clermont* **55**(27), 243 (1975).
- Ducrot, J., Lancelot, J.-R. & Marchand, J. *Earth planet. Sci. Lett.* **62**, 385–394 (1983).
- Peucat, J.-J., Vidal, Ph., Godard, G. & Postaire, B. *Earth planet. Sci. Lett.* **60**, 70–78 (1982).
- Ballard, J.-F., Brun, J.-P. & Durand, J. *C.R. Acad. Sci. Paris* **303**, 1227–1232 (1986).

Evolutionary pressures on planktonic production of atmospheric sulphur

Ken Caldeira

Earth Systems Group, Department of Applied Science,
New York University, 26 Stuyvesant Street,
New York, New York 10003, USA

Dimethylsulphide (DMS), produced by marine phytoplankton^{1–5}, has been proposed as the major source of cloud condensation nuclei (CCN) in the remote marine atmosphere, and as an important factor in global climate¹. Charlson *et al.*¹ suggested that climate modulation and altruism may have been significant factors in the evolution of DMS production by marine phytoplankton. These proposals have led to a re-examination of the relationship between the Earth's biota and climate^{6–11}. Calculations of relative evolutionary pressure in models of individual selection¹² and group selection^{13–17} suggest that neither climate modulation nor altruism could have been the primary factors in the evolution of mid-ocean DMS production. Although a DMS/climate feedback loop may have a role in modulating fluctuations in the Earth's climate^{1,11}, the explanation for mid-ocean DMS production can be found primarily in selection based on local interactions, for example, osmoregulation^{2,5}, and not in evolutionary feedbacks from proposed climate modulation.

Hamilton¹⁸ characterized an altruist as an individual that gives up k units of its own fitness to add K units to the joint fitness of its $(n-1)$ companions. Fitness may be defined for a set of individuals as the average number of offspring surviving to reproduce per individual in that set¹⁹. Here, an altruistic DMS producer is defined as a producer of DMS that decreases its

reproductive rate to increase the reproductive rate of its neighbours. DMS production for any purpose that is directly beneficial to the individual producer would not be characterized as altruistic under this definition, even if it were beneficial to the immediate neighbours of the DMS producer.

Charlson *et al.*¹ suggest that because DMS can be used in osmoregulation and that excess DMS production may have evolved during periods of higher salt stress (for example, glacial epochs), high levels of DMS emission might subsequently be maintained because groups that emit DMS are at a selective advantage. If this possible evolutionary pathway is to be accepted, it is necessary to examine how the altruistic population maintains itself in the face of non-altruistic invaders. Because non-altruistic neighbours of altruists have higher reproductive rates than altruists, altruistic DMS production is not an evolutionarily stable strategy¹²; that is, it will be subject to invasion by non-altruists.

In a schematic model of phytoplankton growth, the growth rate in group i in doublings per day for non-altruistic phytoplankton is $ds_i/dt = P_i$, and for altruists, $da_i/dt = P_i - D$. P_i is the net phytoplankton growth rate without the cost of altruistic DMS production. P_i is a function of local nutrient supply, temperature and other factors. D represents the cost of altruistic DMS production and is a positive value.

A number of workers have investigated models for the maintenance of altruistic populations which assume that the inter-group mixing time is less than the timescale of within-group ecological turnover^{13–16}. The basic idea of these models is that groups containing many altruists grow at a higher rate than groups with few altruists. Even though, in any group, the non-altruists are reproducing more rapidly than the altruists, before the non-altruistic organisms dominate the group, the groups remix and reform a new set of groups. For this form of group

selection to function, it is required that growth rate of altruists exceeds the growth rate of non-altruists:

$$\sum_i \left(\frac{A_i}{\sum_i A_i} \right) \frac{da_i}{dt} > \sum_i \left(\frac{S_i}{\sum_i S_i} \right) \frac{ds_i}{dt} \quad (1)$$

where A_i and S_i are the numbers of altruistic and non-altruistic phytoplankton, respectively, in group i . Inequality (1) must hold consistently and repeatedly, even though, for all i , $da_i/dt < ds_i/dt$.

A mechanism is needed—a trait which results in altruists being grouped with altruists—that produces a distribution of altruists and non-altruists among the groups with supra-binomial variance. Given the high number of density of phytoplankton in the oceans (over 10^6 m^{-3} , ref. 20), it is clear that panmixia will not produce such a distribution. If altruism is to be maintained by group selection, the population is vulnerable to invasion by non-altruistic genotypes that can associate with altruists as if they were themselves altruists. The problem is that any trait which results in being grouped with altruists is favoured by individual selection and thus would not be altruistic¹⁶.

Wright suggested that populations with many altruists might grow faster and emit more migrants than groups of non-altruists, so that groups would tend to receive more altruistic immigrants than non-altruistic immigrants²¹. Even though, in any group the growth rate of non-altruists exceeds that of altruists, it is possible that the altruist growth rate plus the altruist net immigration rate could exceed the non-altruist growth rate plus the non-altruist net immigration rate. Harpending and Rogers¹⁷ analysed a discrete-time mathematical model of group selection based on these ideas. They concluded that altruism is favoured by weak local selection c against altruists, strong group beneficial effects g of altruists, and small local size n . They suggested that altruists can achieve non-zero frequency when

$$n + 1 < \frac{g}{c} \quad (2)$$

What are appropriate values for these variables in the case of mid-ocean DMS production?

Measurements of albedo changes caused by CCN derived from ship stack effluent²² suggest a horizontal scale length of at least 10 km for the diffusion of CCN in the marine atmosphere under stable meteorological conditions. This means that all phytoplankton within a scale area of 100 km^2 would experience essentially the same conditions relating to release of biogenic DMS. If a photosynthetic zone depth of $\sim 100 \text{ m}$ is assumed, and a phytoplankton number density on the order of 10^6 m^{-3} (ref. 20), it is possible to compute a minimal phytoplankton group size of the order of 10^{16} phytoplankters, that is, one phytoplankter's DMS emissions changes the overhead CCN number density less than one part in 10^{16} .

Estimating the cost in the form of doublings per day of DMS production is experimentally problematic. Carbon and sulphur are ubiquitous in the surface oceans, whereas nitrogen, phosphate²³, iron²⁴ or other nutrients may be limiting to phytoplankton growth. Because C, H and S are virtually 'free' in the worlds ocean, the cost of DMS production may be largely the cost of constructing the metabolic machinery that produces DMS molecules. Although the details of this cellular machinery are unknown at this time, and there are conceptual problems associated with evaluating costs of cellular structures with multiple purposes, some idea of the cost of DMS production may be gleaned from comparing DMS carbon output with cellular carbon content. Computations based on data from Vairavamurthy⁵ and Dacey and Wakeham³, using a regression of phytoplankton carbon content as a function of cell size²⁵, indicate a DMS carbon output per day of $\sim 2.7 \times 10^{-4}$ of total cellular carbon (Table 1). Although the cost of producing the cellular machinery that generates DMS is unknown, it is possible to infer, conserva-

Table 1 DMS carbon produced per day as a fraction of total cell carbon (data from refs 3, 5 and 25)

Species	<i>Gymnodinium nelsoni</i>	<i>Hymenomonas carterae</i>
DMS per day (10^{-15} moles)	23 ± 16	1.3 ± 0.1
Cell volume (10^{-12} l cell ⁻¹)	24	0.8
Carbon per cell (10^{-12} moles)	180	9.4
DMS C per day as a fraction of carbon per cell (10^{-4} day ⁻¹)	2.6 ± 1.8	2.8 ± 0.1

tively, a cost of 10^{-6} doublings per day as the cost of producing DMS at today's mean rate.

Using equation 2, with $n = 10^{16}$ and $c = 10^{-6}$, I conclude that for altruism to evolve according to Wright's mechanism²¹, g , the group-benefit effect, would have to be at least of the order of 10^{10} , that is groups composed entirely of altruists would have to emit 10^{10} times as many migrants as groups containing no altruists. This is an implausibly high differential.

If group selection does not maintain planktonic DMS emission, perhaps individual selection does. Let $\partial G/\partial \text{DMS}_{\text{group}}$ be the number of additional doublings per day due to the climatic effects of DMS emission at the mean group DMS emission rate, and let $\partial D/\partial \text{DMS}_{\text{ind}}$ be the cost to the individual in doublings per day of producing DMS at the mean individual DMS emission rate, earlier estimated to be of the order of 10^{-6} . DMS production for climate modulation will be an evolutionarily stable strategy¹² when the cost to the individual of that DMS production is in balance with its direct climatic benefit. Thus, at evolutionary equilibrium, without group selection:

$$\frac{\partial G}{\partial \text{DMS}_{\text{ind}}} = \frac{\partial D}{\partial \text{DMS}_{\text{ind}}} \quad (3)$$

We earlier showed $d \text{DMS}_{\text{group}}/d \text{DMS}_{\text{ind}} < 10^{-16}$ and

$$\frac{\partial G}{\partial \text{DMS}_{\text{ind}}} = \frac{\partial G}{\partial \text{DMS}_{\text{group}}} \frac{d \text{DMS}_{\text{group}}}{d \text{DMS}_{\text{ind}}}$$

This implies that $\partial G/\partial \text{DMS}_{\text{ind}}$ would have to be greater than 10^{10} doublings per day for today's DMS emissions to be explained by individual selection through climatic feedbacks. Maximum measured ocean-phytoplankton doubling rate is on the order of one per day²⁶.

Charlson *et al.*¹ suggest that DMS emitters might be at an advantage because cloud formation with rainfall would return nitrogen to the ocean and also serve as a sunshade. But, if DMS emission increases cloud albedo and diminishes the solar flux to the surface waters, globally, less energy will go into latent heat of evaporation and thus rainfall might actually diminish. Locally, increased CCN number density would lead to smaller cloud droplet size and a diminished likelihood of rainfall, increasing the probability that the clouds would be advected to other regions²⁷. In addition, the nitrogen flux to the mixed layer in rainfall is orders of magnitude smaller than fluxes due to upwelling and mixing of the surface ocean with deeper layers^{23,28}. Although increased cloud albedo would reduce the solar UV flux, it would also reduce the flux in the 400–700 μm band used in photosynthesis, thus moving the compensation depth for phytoplankton growth towards the surface and possibly overwhelming the benefits of reduced UV²⁹. Other factors being equal, up to threshold temperatures, planktonic growth rates are increasing functions of water temperature^{26,30,31}. If there is a reduction in temperature due to DMS emission and increased cloud albedo⁹, phytoplankton growth rates might be reduced.

To balance the high osmotic pressure of seawater (osmolarity $\sim 1.1 \text{ mol l}^{-1}$), phytoplankton need to produce substantial amounts of osmotic and osmoregulatory substances². An efficient osmotic solute must have a relatively low passive

membrane permeability. This permeability depends on the solute's polar character, and therefore molecules with several charged groups will diffuse slowly. Dimethylsulphoniopropionate (DMSP), a DMS precursor, has polar properties that make it highly suitable as osmoregulatory substance. In addition, intra-cellular DMSP concentration varies with salinity in a way consistent with its postulated role as an osmoregulator⁵. Andreae² suggests that a DMSP osmolyte might be selected for in eutrophic regions where the use of a sulphur rather than a nitrogen osmolyte would make all bound nitrogen available for other uses (amino acids, etcetera). The rate of DMS release by phytoplankton increases greatly when the phytoplankton are subjected to grazing by zooplankton, indicating that much DMS release is a result of cell rupture and death³. DMS gas production is 7–26 times higher during the senescence phase of a phytoplankton bloom, when cellular integrity is lost for many phytoplankters, than during the growth phase⁴. This is further evidence that DMS release is a by-product of the need for osmotic regulation. If DMS production is evolutionarily stable, as its ubiquity indicates^{1–3,5}, then the feedback maintaining DMS production is most likely from reasons related to osmolarity and membrane transport, and not through climate modulation.

I thank Tyler Volk for useful discussions, and Alan Grafen and Michael Rampino for helpful comments.

Received 20 September 1988; accepted 3 January 1989

1. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
2. Andreae, M. O. *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 331–362 (Reidel, Dordrecht, 1987).
3. Dacey, J. W. H. & Wakeham, S. G. *Science* **233**, 1314–1316 (1986).
4. Nguyen, B. C. *et al. Mar. Chem.* **24**, 133–141 (1988).
5. Vairavamurthy, E., Andreae, M. O. & Iverson, R. L. *Limnol. Oceanogr.* **30**, 59–70 (1985).
6. Kerr, R. A. *Science* **240**, 393–395 (1988).
7. Lindley, D. *Nature* **332**, 483–484 (1988).
8. Rampino, M. R. & Volk, T. *Nature* **332**, 63–65 (1988).
9. Schwartz, S. E. *Nature* **336**, 441–445 (1988).
10. Legrand, M. R., Delmas, R. J. & Charlson, R. J. *Nature* **334**, 418–420 (1988).
11. Bates, T. S., Charlson, R. J. & Gammon, R. H. *Nature* **329**, 319–321 (1988).
12. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, London, 1982).
13. Charnov, E. L. & Krebs, J. R. *Am. Nat.* **109**, 107–112 (1975).
14. Wilson, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 143–146 (1975).
15. Eshel, I. & Cohen, D. *Population Genetics and Ecology* (eds Karlin, S. & Nevo, E.) 537–546 (Academic, New York, 1976).
16. Nunney, L. *Am. Nat.* **126**, 212–230 (1985).
17. Harpending, H. & Rogers, A. J. *theor. Biol.* **127**, 51–61 (1987).
18. Hamilton, W. D. *Biosocial Anthropology* (ed. Fox, R.) 133–155 (Wiley, New York, 1975).
19. Uyenoyama, M. & Feldman, M. W. *Theor. Pop. Biol.* **17**, 380–414 (1980).
20. Hulbert, E. M. *Limnol. Oceanogr.* **7**, 307–315 (1962).
21. Wright, S. *Ecology* **26**, 415–419 (1945).
22. Coakley, J. A., Bernstein, R. L. & Durkee, P. A. *Science* **237**, 1020–1022 (1987).
23. Broecker, W. S. & Peng, T. H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
24. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
25. Strathmann, R. R. *Limnol. Oceanogr.* **12**, 411–418 (1967).
26. Eppley, R. W. *Fish. Bull.* **70**, 1063–1085 (1972).
27. Rogers, R. R. *A Short Course in Cloud Physics* (Pergamon, Oxford, 1976).
28. Garrels, R. M., Mackenzie, F. T. & Hunt, C. *Chemical Cycles and the Global Environment* (Kaufmann, Los Altos, 1975).
29. Calkins, J. *The Role of Solar Ultraviolet Radiation in Marine Ecosystems* (Plenum, New York, 1982).
30. Goldman, J. C. & Carpenter, E. J. *Limnol. Oceanogr.* **19**, 756–766 (1974).
31. Rhee, G. Y. & Gotham, I. J. *Limnol. Oceanogr.* **26**, 635–648 (1981).

The perception of moving plaids reveals two motion-processing stages

Leslie Welch

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115, USA and University of California at Berkeley, School of Optometry, Minor Hall, Berkeley, California 94720, USA

When viewed through a small aperture, the perceived motion exhibited by a long moving line or grating is ambiguous. This situation prevails because even a perfect machine could only detect motion perpendicular to a moving contour, so motion parallel to a contour is undetectable. The human visual system views the world

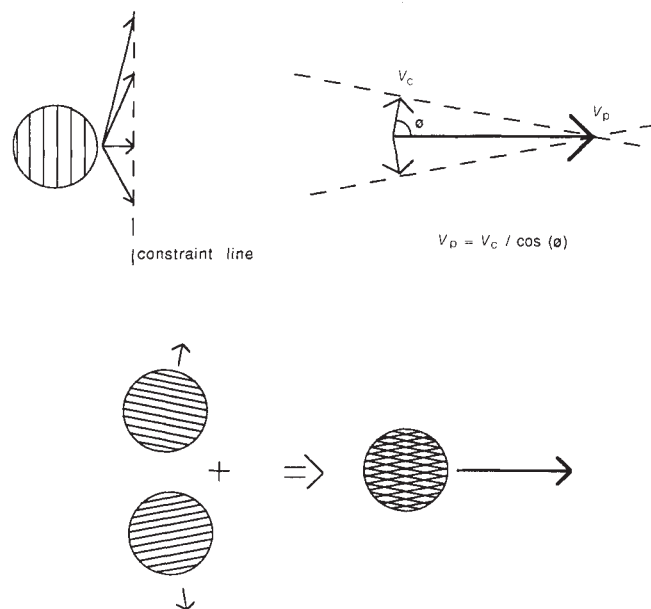


Fig. 1 The motion of a grating viewed through a stationary aperture is ambiguous: motion straight to the right cannot be distinguished from faster motion up and to the right. A family of possible vectors describing the grating motion are constrained to end on the dotted line as illustrated. Two gratings moving in different directions define two constraint lines which intersect at a position designating the velocity that is consistent with the motion of both gratings. The pattern speed is determined by the intersection of constraint lines. Illustrated here are two gratings at $\pm 78^\circ$ forming a plaid. The vector to the intersection of constraint lines indicates that the plaid's speed is much faster than the gratings' speed.

through an aperture array—the neural receptive fields. Therefore a moving object is viewed through many small apertures and the motion within many of those apertures is ambiguous. This ambiguity may be resolved by monitoring the motion of a distinctive feature, such as a line-end or corner, and attributing to the larger object the motion of the feature. Alternatively, Adelson and Movshon¹ have suggested that moving images are processed in two stages, that is, they are first decomposed into one-dimensional components which are later recombined to generate perceived object motion. For a moving plaid, defined as the sum of two drifting gratings (Fig. 1), these alternative models generate different predictions concerning the resolution of the plaid's motion ambiguity. A feature monitor would respond to the motion of the intersections between gratings, whereas the two-stage motion processor would first decompose the plaid into its constituent gratings and subsequently recombine them to generate the perception of a moving plaid. Using speed discrimination to distinguish between the two models, I find that discrimination thresholds reflect the speed of a plaid's component gratings, rather than the speed of the plaid itself. This result supports the two-stage model. Although speed discrimination is limited by component processing, observers cannot directly access component speed. The only perceptually accessible velocity signal is generated by the second-stage pattern processing.

A plaid's speed is equal to the speed of a component grating, in the direction perpendicular to its orientation, divided by the cosine of the angle between the plaid's direction of motion and the grating component's direction of motion. Therefore the plaid always moves faster than the components: the larger the difference in component directions, the greater the difference between the component grating speed and the plaid speed. In this study, the two gratings forming the plaid have different orientations but identical spatial and temporal frequencies. As human speed discrimination varies with the speed of the target, it is appropriate to enquire whether it is the speed of a plaid's

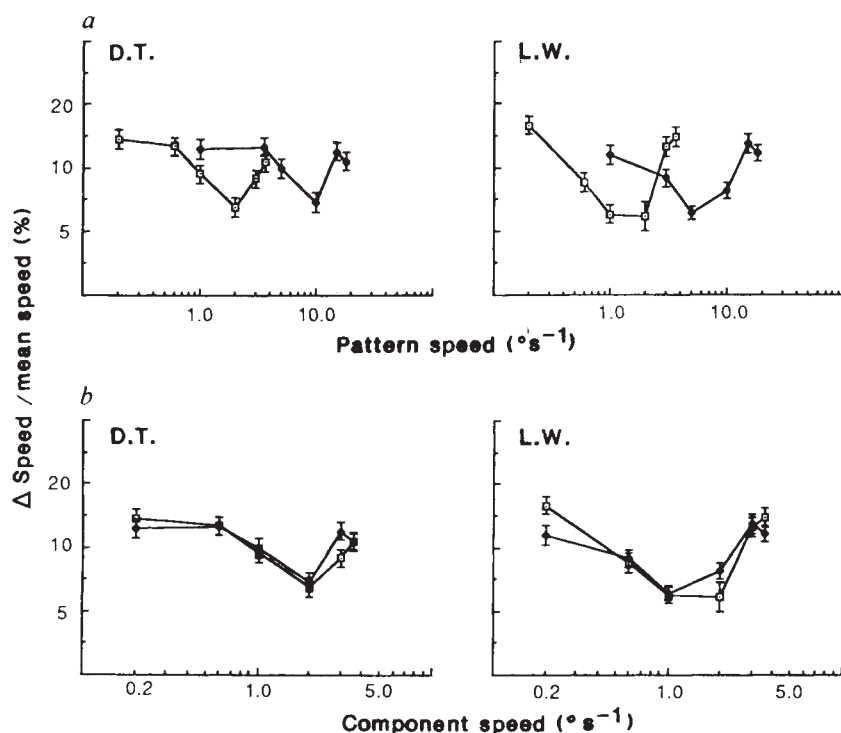


Fig. 2 *a*, Weber fraction percentages for speed discrimination are plotted as a function of pattern speed for two observers, with error bars ± 1 standard error of the mean. The grating contrast was 12.5% and the average luminance 20 cd m^{-2} . The stimuli filled a 4.5° circular aperture and were presented for a variable duration between 0.4 and 0.6 s. Gratings for observers were 3.5 cycles per degree for observer L.W. and 3.0 cycles per degree for observer D.T. Open symbols indicate thresholds for single gratings and closed symbols for plaids. The grating and plaid curves do not superimpose. *b*, Same data as in *a*, replotted as a function of component grating speed. Open symbols signify single gratings and closed symbols signify plaids. The two curves superimpose.

component gratings or the speed of a plaid itself that limits speed discrimination. The feature-monitor model predicts that speed discrimination is limited by the speed of the plaid, whereas the two-stage model predicts that the speed of the component gratings sets the limit.

Plaids formed by gratings at $\pm 78^\circ$ (see Fig. 1) move five times faster than their underlying component gratings. On each trial the stimulus moved at a uniform speed chosen randomly from a narrow range of speeds, and the observers were asked to judge whether it moved faster or slower than the mean speed of the narrow range. Observers (L.W. and D.T.) were provided with practice trials and feedback, and they readily learned the mean speed of the set. The smallest reliably detected change in speed was determined, and that incremental change was divided by the mean speed to calculate the Weber fraction for speed (for more complete methods see ref. 2). Expressed as a percentage, this Weber fraction is the threshold measurement reported here.

Speed discrimination for gratings and plaids are plotted versus pattern speed in Fig. 2*a*. If the pattern speed predicts the thresholds, then identical speed discrimination would be found for a grating moving at 1° s^{-1} and for a plaid moving at 1° s^{-1} . Clearly the predicted relationship does not prevail. Consequently, pattern speed does not predict speed discrimination for plaids. The same discrimination data are replotted as a function of component speed in Fig. 2*b*. If the component speed predicts the thresholds, then discrimination for a grating moving at 1° s^{-1} , would be identical to the value for a plaid moving at 5° s^{-1} , that is, a plaid formed by gratings moving at 1° s^{-1} . This prediction is consistent with my observations. Component speed quantitatively predicts speed discrimination for plaids.

It could be argued that the temporal frequency, namely the rate at which features pass a fixed position in space, is the same for a plaid as for its components, and that the limitation shown in Fig. 2*b* represents a generalized temporal frequency limitation on speed discrimination. Indeed, both temporal frequency and spatial frequency, the spacing between bars of a grating, affect discrimination. To study the effects of temporal frequency, I generated a plaid pattern made of high-spatial-frequency components: 12 cycles per degree for observer L.W. and 15 cycles per degree for observer D.T., but moderate temporal frequency, 6 Hz. Discrimination for the high-spatial-frequency plaid was

compared with a grating characterized by the same temporal frequency and moving with the same speed as the plaid. It should be noted that the spacing of the dark bars of the control grating is identical to the spacing of the plaid intersections along any given horizontal section. Figure 3 displays the following results: discrimination for the plaid was poor, but discrimination for the grating with the identical temporal frequency was excellent. Clearly, the pattern's temporal frequency does not limit speed discrimination.

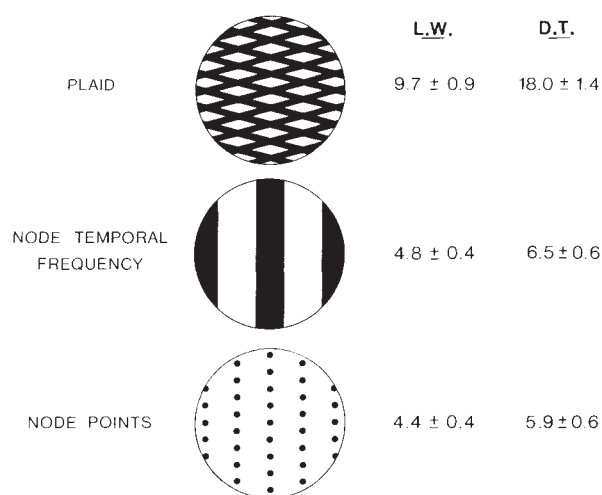


Fig. 3 Speed-discrimination thresholds, expressed as percentage of mean speed, are compared for three patterns moving with the same speed. For observer L.W., all patterns moved at 2.5° s^{-1} . The plaid was the sum of 12 cycles per degree gratings drifting at 6 Hz, and the node temporal frequency grating was 2.4 cycles per degree drifting at 6 Hz. For observer D.T., the patterns moved at 2.0° s^{-1} ; the plaid was the sum of 15 cycles per degree gratings drifting at 6 Hz, and the node temporal frequency grating was 3.0 cycles per degree drifting at 6 Hz. Note that the spatial frequency of the 'node temporal frequency' grating is the same as the spacing between the nodes for a single horizontal section. Thresholds for the plaids are higher than the thresholds for either of the other two patterns.

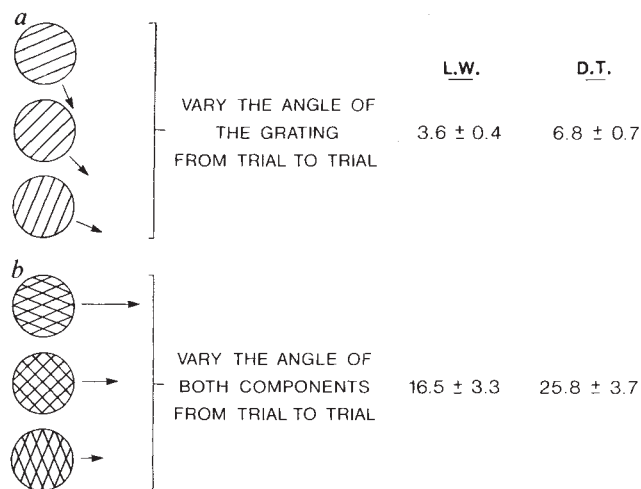


Fig. 4 *a*, A single grating at various angles; angles were chosen randomly on each trial from a set of 7 fixed angles ranging from 62° to 28°, or 45° mean. Gratings were 2 cycles per degree spatial frequency, and 5 Hz temporal frequency. Speed discrimination for this condition was good. The lower spatial frequency of these gratings results in slightly better speed discrimination thresholds than the experiment shown in Fig. 2. *b*, A plaid with both components changing angles; angles were again between 62° and 28°. The high discrimination thresholds indicate that this task was difficult. A constant-angle plaid results in thresholds identical to those found for the single grating in *a*.

The salient features of a plaid are the intersections of the component gratings which form lighter and darker 'nodes.' If the nodes limit speed discrimination for a plaid, a pattern made up of the nodes alone should produce a threshold value equal to that for the plaid. To test this idea, I used a grid of dots with spatial positions and speed identical to the dark nodes (illustrated at the bottom of Fig. 3). Speed discrimination for the dot-grid was superior to the thresholds for the corresponding plaid (Fig. 3).

Both the present data and plaid-masking data³ support the Adelson and Movshon two-stage model^{1,4}. These results also demonstrate that the component stage is the site of limiting noise for precise speed discrimination. This does not imply that the signal at the component stage is accessible to the observer for speed discrimination. Rather, the information must pass through the second or pattern-processing stage before it becomes available to the observer. The following experiment demonstrates this constraint: discrimination for single gratings, with orientations randomized between trials, is compared with discrimination for plaids, whose components were similarly varied. Speed discrimination for a single grating was not impaired by randomized orientation, as shown in Fig. 4. Random orientation of the single gratings does not affect component speed, but varying the angle between the components of a plaid pattern introduces large variations in the pattern speed. By this procedure, the speeds of the components and the plaid are decoupled. Hence, on any single trial, when the components moved faster than the mean speed of the test set, the plaid may actually have moved slower than the mean, because the angle between the components was decreased on that particular trial. During these trials, observers were asked to discriminate the speed of the plaid's component gratings. The data in Fig. 4 indicate that the observers viewing the plaid were unable to judge the component's speed accurately, although the same information was available here as in the single grating instance. In conclusion, it follows that the pattern speed is the only information to which the observers could respond.

The Adelson and Movshon two-stage model receives further support from our data; component processing occurs before pattern processing, with no speed information from the com-

ponent stage by-passing the pattern stage. First, a complex pattern is decomposed into its component parts and later these are combined to form a coherent moving pattern. The limiting noise is localized in the component-processing stage, which is followed by the pattern-processing stage where a velocity vector is extracted. In sum, the observer perceives the velocity of a complex pattern with a precision limited by the component-processing stage. Nonetheless, the velocity signal is available to perception only after the pattern-processing stage.

I thank Dr Ken Nakayama for suggesting the original experiment and for discussion and Drs Suzanne McKee and Samuel Bowne for help and discussion. This work was supported by the US Air Force Office of Scientific Research and the National Eye Institute.

Received 1 August 1988; accepted 9 January 1989.

- Adelson, E. H. & Movshon, J. A. *Nature* **300**, 523-525 (1982).
- McKee, S. P., Silverman, G. H. & Nakayama, K. *Vision Res.* **26**, 609-619 (1986).
- Ferrera, V. P. & Wilson, H. R. *Vision Res.* **27**, 1783-1796 (1987).
- Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. *Pattern Recognition Mechanisms* (Springer, New York, 1985).

Pathological changes in olfactory neurons in patients with Alzheimer's disease

Barbara R. Talamo, RuthAnn Rudel, Kenneth S. Kosik*, Virginia M.-Y. Lee†, Samuel Neff, Lester Adelman & John S. Kauer

Neurosciences Laboratories and Departments of Neurology, Neurosurgery and Pathology, Tufts Medical School and New England Medical Center, Boston, Massachusetts 02111, USA
* Center for Neurologic Diseases, Harvard Medical School, Boston, Massachusetts 02115, USA

† Department of Pathology and Laboratory of Medicine, University of Pennsylvania Medical School, Philadelphia, Pennsylvania 19104, USA

Alzheimer's disease is a central nervous system disorder characterized by the presence of neurofibrillary tangles, neuritic plaques and dystrophic neurites¹ in susceptible areas of the brain. Investigation of the mechanism and development of the disease has been hampered by the lack of an animal model and the inaccessibility of neural tissue during the illness. Deficits in odour detection and discrimination are among the signs of Alzheimer's² and previous anatomical studies suggest that olfactory pathways may be involved early in the illness³. Neurons in the olfactory epithelium, which are of central origin, are relatively accessible for biopsy⁴ and could be used as a source of living nerve cells for the study of Alzheimer's disease if they can be shown to have characteristics of this disease. As these neurons have the unusual property of arising from stem cells throughout the life of the organism⁵, they are good candidates for the development of cell cultures or cell lines which may express the disorder from living patients. We report here that nasal epithelium tissue taken at autopsy shows unique pathological changes in morphology, distribution and immunoreactivity of neuronal structures in patients with Alzheimer's disease.

In the normal human adult, olfactory receptor neurons are distributed in a patchy way. Sensory regions were identified by immunohistochemical staining for the olfactory neuro-enriched protein olfactory marker protein (OMP)⁶ (Fig. 1a) or for human neuron-specific enolase (NSE; data not shown). In olfactory neurons, only axons were stained with antibodies to neurofilament (NF) proteins, and these seemed to contain only one of the NF subunits. These axons reacted strongly with a monoclonal antibody specific to the dephosphorylated form of the mid-sized neurofilament subunit, NFM(P-) (antibody RMD020, Table 1). They did not react, however, with mono-

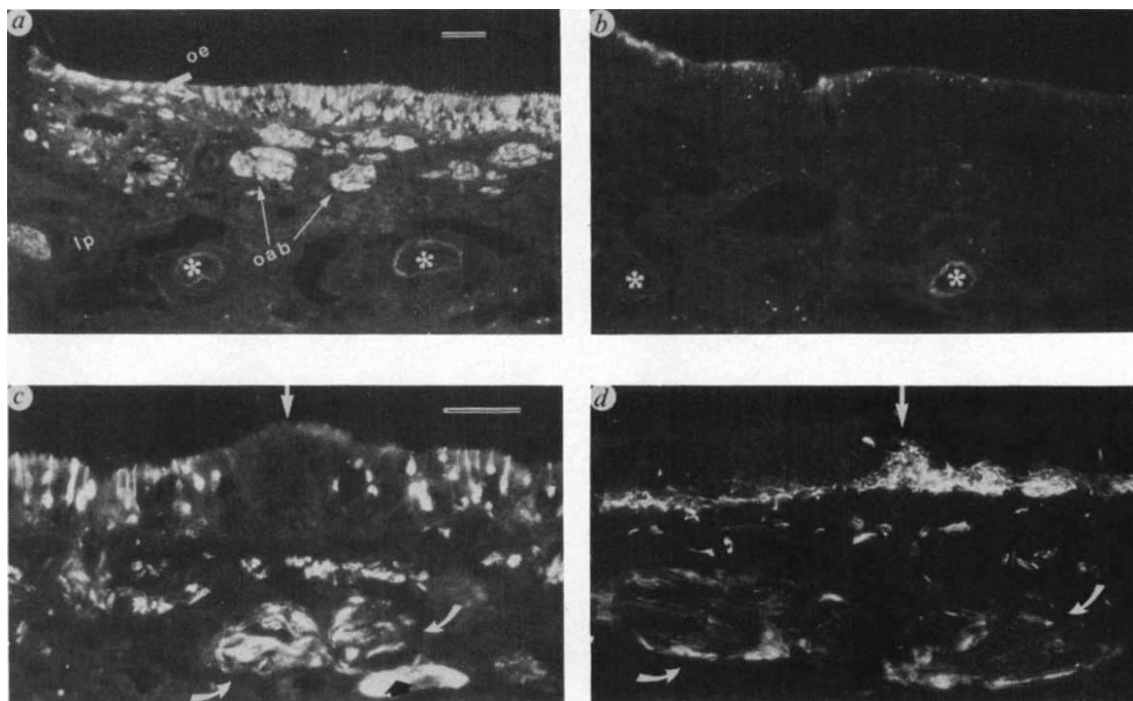


Fig. 1 *a*, Non-AD (patient A87.121) Olfactory epithelium stained with OMP antibody. Olfactory receptor cells and olfactory nerve bundles are positively stained. Asterisks label blood vessels, ringed with an autofluorescent muscle layer, that are common to *a* and *b*. *oe*, Olfactory epithelium; *oab*, olfactory axon bundle; *lp*, lamina propria. Scale bar, 100 μ m. *b*, Corresponding non-AD section stained with antibody to NFH(P+++). There is no positive staining in this section. Autofluorescence along the epithelial surface is due to pigment granules. Magnification is as in *a*. *c*, AD olfactory epithelium (patient A87.30) stained with antibody to OMP. Olfactory receptor cells and axons are stained. An epithelial region devoid of receptor cells is marked with a vertical arrow. Axonal bundles common to this section and corresponding section (*d*) are marked with curved arrows. Autofluorescent pigment is noted with a black arrowhead. Scale bar, 100 μ m. *d*, Corresponding AD section stained with antibody to NFH(P+++). Neurites stained for NFH(P+++), project densely below the basement membrane and up into the olfactory epithelium which has no neurons at that point (vertical arrow in *c* and *d*). Olfactory axon bundles also show staining for NFH(P+++). Magnification, as in *c*.

Methods. Olfactory epithelium was removed at autopsy by undercutting the cribriform plate and removing it as a unit. Olfactory epithelium and bulbs were placed in 2% paraformaldehyde/10% saturated picric acid in 0.1 M sodium phosphate buffer, pH 7.4 (PB), usually for 2–4 h (fixation for up to 2 days did not seem to affect the result). After dissecting away cartilage and bony material, the tissue was rinsed in cold PB, placed in 30% sucrose/PB at 4 °C overnight and frozen. Storage as a block or as 8–10 μ m frozen sections at –70 °C for up to 18 months produced no changes in immunoreactivity. Sections were treated with 3% blocking serum (same species as secondary antibody)/0.1% Triton X-100 in PB for 20–30 min before incubating with primary antibody, usually for 2 h at room temperature or overnight at 4 °C. Using standard methods, the appropriate biotinylated secondary antibody was applied in blocking serum followed by treatment with streptavidin-Texas Red (1:350 in 2% BSA; BRL). Solutions were buffered in PB except for monoclonal antibodies to NF and also ALZ50, which were buffered in Tris. Slides were blocked with 5% non-fatty dry milk for ALZ50. Antibody reactivity was monitored by including known positive control olfactory epithelium or bulb in each run. AD cortex fixed in formalin was the control for ALZ50 staining. In other controls primary antiserum was omitted. Mouse monoclonal antibodies (dilution) used: RM024 (none); RM0254 (1:10); RMd020 (1:10); glial fibrillary acidic protein (1:10) and rat monoclonal H014 (1:10)^{29,30}; tau (5E2) (1:10)¹⁵; vimentin, clone V-9 (1:1,000) (ICN); ALZ50 (1:5) (P. Davies); anti-NF (1:200) (Labsystems). Polyclonal rabbit antibodies to: NFL (1:500) (V. Lee); cyokeratin (1:300) (DAKO); human NSE (1:2,500) (Polysciences); intermediate filament protein 57K (1:300) (L. Parysek); and goat antibody to OMP (1:200 to 1:1,000) (F. Margolis). Biotinylated secondary antibodies: rabbit anti-goat IgG, horse anti-mouse IgG and goat anti-rabbit IgG (Vector); goat anti-mouse IgM and anti-rat IgG (Tago). The alkaline phosphatase treatment³¹ used 20 U ml⁻¹ enzyme and ZnCl₂ instead of ZnSO₄. Control sections were treated in the same way but with no enzyme.

clonal antibodies directed against other forms of NFM or against various epitopes of heavy (NFH) or light (NFL) subunits. For example, there was no evidence that axonal NFM contained phosphorylated—NFM(P+)—or phosphate-independent—NFM(Pind)—epitopes, or phosphorylated (Fig. 1*b*) or dephosphorylated epitopes of NFH or NFL (Table 1). The perikarya and dendrites of olfactory neurons did not react with any of these antibodies. As far as we know, this is the best evidence for the presence of any NF subunit in olfactory neurons in normal adult animals. Other intermediate filament proteins, including vimentin (present in rat olfactory neurons⁷), glial fibrillary acidic protein, the 57K protein⁸ and cyokeratin are not detectable. The neuronal microtubule-associated protein tau was not detectable with monoclonal antibody (5E2) in any part of the olfactory receptor neuron in normal tissue under conditions that label Alzheimer's brain tissues.

Autopsied nasal epithelium from patients with Alzheimer's disease (AD) contained both sensory and non-sensory areas

(Fig. 1*c*) but, in contrast to control tissue, had increased reactivity to neurofilament antibodies and abnormal neuronal structures. In particular, olfactory nerve axon bundles (identified by their location and by staining for OMP, also reacted with antibodies to NSE and NFM(P–) and were strongly immunoreactive for phosphorylated forms of NFH (Fig. 1*d*), phosphate-independent forms of NFM, and for NFL (data not shown, see also Table 1). Extensive neurite accumulations were also observed in some sensory regions above and just below the epithelial basal lamina and projecting into the epithelium (Fig. 2). The masses of immunoreactive fibres often displaced epithelial cells (Fig. 1*d*, arrow). These unusual neurites reacted positively with NSE antibody, with all the NF antibodies that stained axons, with antibodies to NFH(P–), NFM(P+) and, most strongly, with antibodies to NFH(P+++), (see Table 1), to another NF antibody (Labsystems) that recognized a phosphatase-sensitive epitope (Table 1), and also to NFM(Pind). Increased phosphorylation of cytoskeletal proteins has also been

Table 1 Specificity and staining localization of antibodies

Antibody	Specificity	Perikaryon		Axons		Neurites	
		Control	ALZ	Control	ALZ	ALZ only	ALZ only
					Phosphatase treated	Phosphatase treated	
mcl RM024	NFH (P+++)	—	—	—	+	—	+
mcl TA51	NFH (P+)	—	+/-	—	+	ND	+
mcl DP1	NFH > NFM (P-)	—	—	—	—	ND	+
mcl HO14	NFM (P+)	—	—	—	—	ND	+
mcl RMO254	NFM (Pind)	—	—	—	+	+	+
mcl RMD020	NFM > NFH (P-)	—	—	+	+	+	+
pcl NFL	NFL	—	—	—	+	ND	+
pcl anti-NF (Labsys)	? (P+)	—	—	—	+	—	+
mcl ALZ50		—	—	—	—	ND	+
pcl OMP		+	+	+	+	+	—
pcl NSE		+	+	+	+	ND	+
mcl GFAP		—	—	—	—	ND	—
mcl Vimentin		—	—	—	—	ND	—
pcl p57	57K intermediate filament	—	—	—	—	ND	—
mcl 5E2	Tau (Pind)	—	—	—	—	ND	+

Phosphatase treatment was as in the Fig. 1 legend. Antibody specificities were established by western blot (data not shown).

(P+++), heavily phosphorylation-dependent; (P+), phosphorylation-dependent; (Pind), phosphorylation-independent; (P-), dephosphorylation-dependent. ND, not done; mcl, monoclonal; pcl, polyclonal; +, present; —, absent.

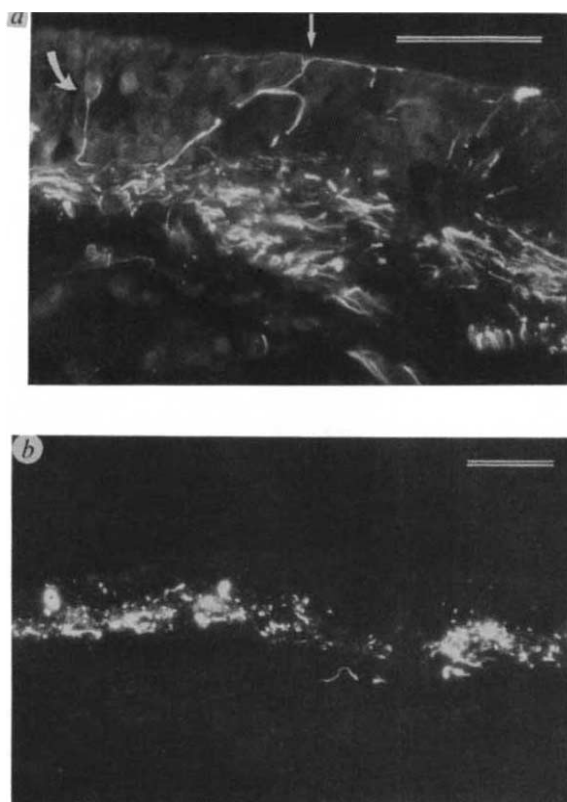


Fig. 2 AD olfactory epithelium stained with antibody to NF (Labsystems anti-NF; patient A87.30). Single neurites positively stained for NF project anomalously into the olfactory epithelium. Arrow points to nerve fibre invading the epithelium and running along the epithelial surface. A curved arrow points to what seems to be a rarely-stained olfactory neuron and its axon. The control epithelium looks similar to that in Fig. 1b. Calibration bar, 100 µm. b, AD olfactory epithelium stained with antibody ALZ50 (patient A87.30). ALZ50 stains the neuritic layer below the epithelium, but staining is less extensive than for NFH(P+). Staining is confined to short fibres just below the basement membrane and globular structures within the epithelium. Deep olfactory nerve bundles are unstained. No positive staining is observed in non-AD epithelium. Calibration bar, 100 µm.

reported in affected neurons in AD⁹⁻¹². No neurofibrillary tangles were observed in AD olfactory epithelium.

Cortical and hippocampal brain regions were examined for AD characteristics. Table 2 shows a comparison between levels of plaque formation and neurofibrillary tangles in the hippocampus and levels of neurite accumulation detected by NFH(P+) immunoreactivity in the olfactory epithelium. Neurite accumulation in the olfactory epithelium correlated with AD-type brain pathology. Of 14 control patients, only two had NF-positive neurites in the olfactory epithelium. Of nine AD patients, evidence of light to extensive NFH(P+) positive accumulations of neurites was found in eight cases. The inconsistent results in the two control cases and the single AD case remain unexplained. There was no evidence of dementia in the two 'control' patients. Patient A86.127 had no obvious pathology in the hippocampus or temporal lobe; however, numerous plaques and tangles were present in the frontal cortex. Patient A88.15 did not have plaques and tangles either, but did have a substantial loss of neurons in hippocampal region CA-1. The demented patient XE 86.48 showed severe wasting and brain atrophy, had OMP-positive olfactory receptor neurons in her epithelium, but there was no evidence of staining for phosphorylated NF. Patients with other neurological symptoms did not have disturbances of the olfactory epithelium, although the sample is small and these patients were younger, on average, than the AD patients.

Neuritic masses in AD olfactory epithelium were also stained in some cases with antibody ALZ50, which was reported to be completely specific to Alzheimer's tissue¹³ (Fig. 3), and with a monoclonal antibody to tau protein (5E2, data not shown); both antibodies stain neurofibrillary tangles and neuritic plaques in AD brains¹³⁻¹⁵. ALZ50, which cross-reacts with tau¹⁶, had a similar staining pattern to 5E2 near the basal lamina. Deep axon bundles of olfactory neurons did not stain with these antibodies. Tau is the major antigenic epitope of neurofibrillary tangles^{9,17-20}; the issue of whether NF proteins are also present in tangles^{14,21} is controversial because many NF antibodies cross-react with tau. It seems that both NF and tau expression are altered in AD olfactory epithelium as the NF antibodies used in our study do not cross-react with tau²², nor do they stain tangles (except for RMD020). NF subunit expression and phosphorylation were increased in both axons and the neuritic network at the basal lamina, whereas tau was present mainly in the neurites and in some unidentified globular structures in the

Table 2 Comparison of brain pathology with staining for phosphorylated neurofilament in the olfactory epithelium

Epithelium	Age/sex	Primary diagnosis	CA-1, hippocampus		Olfactory epithelium phosphorylated neurofilament staining
			Plaques	Tangles	
A86.120	61/M	Myocardial infarction	—	—	—
A87.6	71/F	Cardiogenic shock	+	—	—
A87.32	67/F	Hepatic failure	++	—	—
A87.108	79/M	Diabetes, cancer, atherosclerosis	—	+	—
A87.121	80/F	Stroke, seizures	—	—	—
B1174	94/M	Cancer	—	+	+/-
B885	41/F	Depression/suicide	—	—	—
A87.54	45/M	Puncture wounds	—	—	—
A88.21	29/M	Astrocytoma, cortex, hippocampus, medulla reactive gliosis in cortex	—	—	—
A88.11	36/F	Metachromatic leukodystrophy	—	—	—
A88.45	29/F	Retardation, seizures, cancer	—	—	—
B1217	59/F	Huntington's disease	—	—	—
A86.127	72/M	Cancer	—	—	+++
A88.15	83/M	Cardiovascular disease; cancer	—	—	++
XE86.50	84/M	AD	+	+++	++
XE87.10	77/F	AD, seizures	++	+	+
A87.30	91/F	AD	+	+++	+++
B1152	82/M	AD	++	++	+++
XE88.1	84/M	AD	++	++	+++
B1204	81/F	AD	++	++	+++
B1276	92/F	AD	+++	+++	+++
XE88.27	89/F	AD	+++	+++	+/-
XE86.48	~70/F	AD	+++	++	—

Patients that were considered to have Alzheimer's disease had been clinically diagnosed as such and had numerous senile plaques on Bielschowsky impregnation in hippocampus, temporal lobe, neocortex and frontal cortex. In all three areas the counts exceeded those required for a diagnosis of AD²⁸. Neurofibrillary tangles were numerous in the hippocampus in every case, and were found in variable numbers in neocortex. There was no anatomic evidence of infarction or any other disease associated with dementia in AD patients. One group of control patients consisted of age-matched cases with no history of dementia and with only an occasional neurofibrillary tangle or senile plaque. Other controls had retardation or dementia clearly associated with non-AD pathology. Paraffin sections (8 µm) of the hippocampal region of formalin-fixed tissue were examined. Plaques and tangles in the CA1 region were counted in a 100× field. Scoring of plaques and tangles used in this table was as follows. (—), fewer than five plaques and one or fewer tangles; (+), five to nine plaques, two to ten tangles; (++) , 10–40 plaques, 11–20 tangles; (+++) , more than 40 plaques or more than 20 tangles. Neuritic staining with antibody to phosphorylated neurofilament was scored as: rare (+/-), or intermittent (+), to heavy and extensive (+++). The interval between death and tissue fixation seemed to have no effect on immunocytochemical detection of neurofilaments. The range was similar for control cases (4–32 h; mean = 15.7 ± 7.9 h (s.d.)) and AD cases (4.5–26 h; mean = 16.0 ± 6.7 h).

epithelium (Fig. 3). These unusual neurite masses in the epithelium may be analogous to the neurites around plaques or dystrophic cortical neuropil structures in AD brain, all of which stain for tau¹⁵. The reactivity of the cortical structures with antibodies that are completely specific to neurofilament is not known.

The neuritic accumulation under the epithelium and the increase in tau-immunoreactive neuropil fibres in AD cortex may reflect an abnormally-induced proliferative response²³. The continued turnover of olfactory receptor cells in adult animals and their lack of the usual complement of NF proteins indicates that they retain features of immature neurons²⁴ and thus may be more responsive to trophic signals than adult CNS neurons are, although it is not clear whether the neurites are derived from olfactory receptor cells.

Other demented patients who show olfactory deficits and structures similar to neurofibrillary tangles (for example some Parkinson's and Down's patients)^{25–27} may have similar pathology. Biopsy studies are in progress to examine the level of olfactory epithelial pathology in patients with dementia or other neurodegenerative diseases and to correlate these results with clinical measures of dementia and olfactory deficit.

We thank Dr Seymour Reichlin for his support and suggestions, Dr Edward D. Bird and the Brain Tissue Resource Center for making human brain and olfactory tissue available (supported in part by the PHS and the Hereditary Disease Foundation), Dr Ed Stopa for helpful discussions, Dr Richard Shader for pointing out the presence of olfactory deficits in putative Alzheimer's patients, and Walter Dent for technical assistance. This project is supported by grants from the Pew Freedom Trust and the NIH (V.M.-Y.L.).

Received 9 October 1988; accepted 10 January 1989.

- Braak, H., Braak, E., Grundke-Iqbal, I. & Iqbal, K. *Neurosci. Lett.* **65**, 351–355 (1986).
- Serby, M., Corwin, J., Conrad, P. & Rotrosen, J. *Am. J. Psychiat.* **142**, 781–782 (1985).
- Pearson, R. C. A., Esiri, M. M., Hiorns, R. W., Wilcock, G. K. & Powell, T. P. S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4531–4534 (1985).
- Lovell, M. A., Jafek, B. W., Moran, D. T. & Rowley, J. C. III. *Arch. Otol.* **108**, 247–249 (1982).
- Graziadei, P. P. C., Karlan, M. S., Monti-Graziadei, G. A. & Bernstein, J. J. *Brain Res.* **186**, 289–300 (1980).
- Margolis, F. L. in *Advances in Neurochemistry*, 193–246 (Plenum New York, 1975).
- Schwob, J. E., Farber, N. B. & Gottlieb, D. T. *J. Neurosci.* **6**, 208–217 (1986).
- Parysek, L. M. & Goldman, R. D. *J. Neurosci.* **8**, 555–563 (1988).
- Grundke-Iqbal, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4913–4917 (1986).
- Haugh, M. C., Probst, A., Ulrich, J., Kahn, J. & Anderton, B. H. *J. Neurol. Neurosurg. Psychiat.* **49**, 1213–1220 (1986).
- Sternberger, N. H., Sternberger, L. A. & Ulrich, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4274–4276 (1985).
- Cork, L. C. *et al. J. Neuropath. exp. Neurol.* **45**, 56–64 (1986).
- Wolozin, B. L., Pruchnicki, A., Dickson, D. W. & Davies, P. *Science* **232**, 648–650 (1986).
- Nukina, N., Kosik, K. S. & Selkoe, D. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3415–3419 (1987).
- Kowall, N. W. & Kosik, K. S. *Ann. Neurol.* **22**, 639–643 (1987).
- Ksiezak-Reding, H., Davies, P. & Yen, S.-H. *J. biol. Chem.* **263**, 7943–7947 (1988).
- Wood, J. G., Mirra, S. S., Pollock, N. J. & Binder, L. I. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4040–4043 (1986).
- Kosik, K. S., Joachim, C. L. & Selkoe, D. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4044–4048 (1986).
- Wischik, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4884–4888 (1988).
- Brion, J. P., Passareiro, H., Nunez, J. & Flament-Durand, J. *Archs Biol. Brux.* **95**, 229–235 (1985).
- Ksiezak-Reding, H., Dickson, D. W., Davies, P. & Yen, S.-H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3410–3414 (1987).
- Lee, V. M.-Y., Otvos, L., Jr., Schmidt, M. L. & Trojanowski, J. Q. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7384–7388 (1988).
- Carrell, R. W. *Nature* **331**, 478–479 (1988).
- Carden, M. J., Trojanowski, J. Q., Schlaepfer, W. W. & Lee, V. M.-Y. *J. Neurosci.* **7**, 3489–3504 (1987).
- Ansari, K. A. & Johnson, A. *J. chron. Dis.* **28**, 493–497 (1975).
- Warner, M. D., Peabody, C. A. & Berger, P. A. *Soc. Biol. Psych.* **23**, 836–839 (1988).
- Mann, M. & Esiri, M. M. *New Engl. J. Med.* **318**, 789–790 (1988).
- Khachaturian, Z. S. *Arch. Neurol.* **42**, 1097–1105 (1985).
- Lee, V. M.-Y., Carden, M. J., Schlaepfer, W. W. & Trojanowski, J. Q. *J. Neurosci.* **7**, 3474–3488 (1987).
- Schmidt, M. L., Carden, M. J., Lee, V. M.-Y. & Trojanowski, J. Q. *Lab. Invest.* **56**, 282–294 (1987).
- Lee, V. M.-Y., Carden, M. J. & Trojanowski, J. Q. *J. Neurosci.* **6**, 850–858 (1986).

Extrusion of calcium from rod outer segments is driven by both sodium and potassium gradients

L. Cervetto*, L. Lagnado, R. J. Perry, D. W. Robinson & P. A. McNaughton

Physiological Laboratory, Downing Street, Cambridge CB2 3EG, UK and * Istituto di Neurofisiologia del CNR, via San Zeno 51, 56100 Pisa, Italy

Calcium is transported across the surface membrane of both nerve^{1,2} and muscle³ by a Na⁺-dependent mechanism, usually termed the Na:Ca exchange. It is well established from experiments on rod outer segments that one net positive charge enters the cell for every Ca²⁺ ion extruded by the exchange⁴⁻⁷, which is generally interpreted to imply an exchange stoichiometry of 3 Na⁺:1 Ca²⁺. We have measured the currents associated with the operation of the exchange in both forward and reversed modes in isolated rod outer segments and we find that the reversed mode, in which Ca²⁺ enters the cell in exchange for Na⁺, depends strongly on the presence of external K⁺. The ability of changes in external K⁺ concentration ([K⁺]_o) to perturb the equilibrium level of [Ca²⁺]_i indicates that K⁺ is co-transported with calcium. From an examination of the relative changes of [Ca²⁺]_o, [Na⁺]_o, [K⁺]_o and membrane potential required to maintain the exchange at equilibrium, we conclude that the exchange stoichiometry is 4 Na⁺:1 Ca²⁺, 1 K⁺ and we propose that the exchange should be renamed the Na:Ca, K exchange. Harnessing the outward K⁺ gradient should allow the exchange to maintain a Ca²⁺ efflux down to levels of internal [Ca²⁺] that are considerably lower than would be possible with a 3 Na⁺:1 Ca²⁺ exchange.

The activity of the electrogenic exchange was measured from the membrane current produced by its operation in an isolated rod outer segment under whole-cell voltage clamp⁷. Two advantages of this preparation are that the exchange is the only significant carrier of membrane current when light-sensitive channels have been suppressed⁷, and that transmembrane fluxes mediated by the exchange far outweigh other calcium fluxes, either across the membrane or into the whole-cell pipette. The time constant of equilibration of calcium across the membrane is of the order of 10 seconds or less when the exchange is active under the conditions of the present experiments, whereas it exceeds 100 seconds when operation of the exchange is prevented (see Fig. 4 legend).

Figure 1a shows that reducing [Ca²⁺]_o activates a transient inward current as calcium is extruded by the exchange⁴⁻⁷ and that restoring [Ca²⁺]_o activates a transient outward exchange current as the previous equilibrium level of [Ca²⁺]_i is re-established. The integrals of the inward and outward currents are equal (Fig. 1b), showing that the charge moved by the exchange per Ca²⁺ ion transported is equal in both forward and reversed modes of the exchange.

The reversed mode depends on the presence of external K⁺ (Fig. 2). Reducing external [Na⁺] to zero in the presence of external K⁺ activates a transient outward current (trace 1) which is associated with a Ca²⁺ influx, because return to normal [Na⁺]_o activates an inward exchange current. Repeating the experiment in the absence of external K⁺ produces a much smaller inward current (trace 2), which is probably due to a small shift in membrane potential caused by the junction potential between the Na⁺-containing and Na⁺-free solution (see Fig. 2 legend). The difference between traces 1 and 2 is shown in Fig. 2b.

Figure 2 demonstrates that the reversed exchange depends on external K⁺, but does not show whether K⁺ is co-transported or is simply a co-factor. Changing the K⁺ gradient across the cell membrane will perturb the equilibrium level of [Ca²⁺]_i if K⁺ is co-transported, whereas if it is a co-factor, forward and

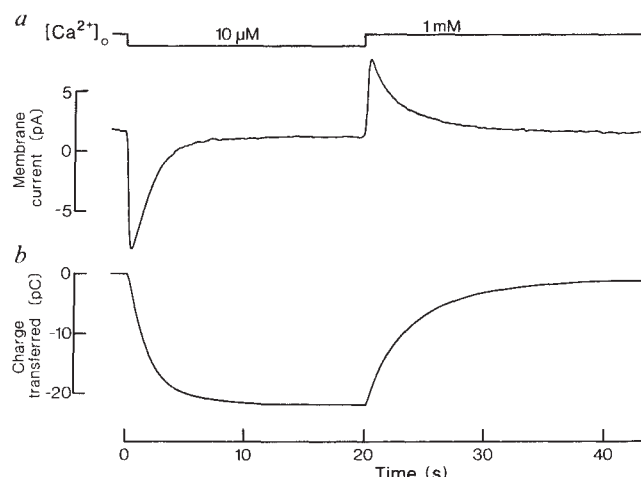


Fig. 1 *a*, Effect on the membrane current, recorded from a rod outer segment isolated from the retina of the salamander *Ambystoma tigrinum*, of changing the solution from Ringer, with 1 mM [Ca²⁺]_o, to a solution containing 10 μM [Ca²⁺]_o. Similar results were observed in 10 experiments. *b*, Integral of the transient changes in current. Total charge flowing during exposure to 10 μM [Ca²⁺]_o was -21.8 pC, and the charge flowing on readmission of 1 mM [Ca²⁺]_o was 21.0 pC, corresponding to a mean change in total [Ca²⁺]_i of 222 μM in a cell volume of 10⁻¹² litres. The predicted change in free [Ca²⁺]_i for a 4 Na⁺:1 Ca²⁺, 1 K⁺ exchange at -14 mV is 34 μM (see equation 1), and the difference between this and the measured change in total [Ca²⁺]_i is satisfactorily accounted for by the ratio of about 1:20 between free and total [Ca²⁺]_i (see ref. 5). For a 3 Na⁺:1 Ca²⁺ exchange the predicted change in free [Ca²⁺]_i would be 427 μM, which is inconsistent with the smaller change in total [Ca²⁺]_i.

Methods. Rod outer segment maintained in Ringer's solution, composition (in mM): NaCl, 110; KCl, 2.5; MgCl₂, 1.6; CaCl₂, 1; HEPES 10, neutralized to pH 7.6 with tetramethylammonium hydroxide, and exposed to 110 mM NaCl, 0.27 mM CaCl₂, 2 mM nitrilotriacetic acid (10 μM free Ca²⁺), 5 mM KCl, 10 mM HEPES, pH 7.6 for 20 s (duration of exposure is shown at top). The solution was changed using an automated multi-pipe stepping system¹⁴, in which the solution bathing the outer segment is changed completely in 50 ms. Outer segments in this and other figures were voltage-clamped at -14 mV using the whole-cell method^{7,20}. The whole-cell pipette contained (in mM): sodium aspartate, 100; potassium aspartate, 10; MgCl₂, 3; Na₂ATP, 1; PIPES, 10 neutralized to pH 7.0 with KOH (total [K⁺] = 28.5 mM); note that [Na⁺] was high to activate the reversed mode of the exchange, and that cGMP and GTP were omitted to ensure that light-sensitive channels were closed. The exchange did not depend directly on the presence of ATP as similar results were obtained in the absence of both ATP and MgCl₂. The currents observed in these experiments are due to the operation of a Na⁺-dependent calcium transport mechanism and not to the passage of ions through membrane channels because (1) inward current is promoted by removal of calcium or potassium, ruling out a direct flux of Ca²⁺ or K⁺ through channels; (2) no increase in noise is observed during current activation; (3) the time constant of relaxation of current depends strongly on the concentrations of Na⁺, Ca²⁺ and K⁺ on both sides of the membrane, and only weakly on membrane potential (see Fig. 4a); and (4) the charge transfer caused by a change in ionic conditions or membrane potential is always equal, within experimental error, to that on return to control conditions. The currents are not due to operation of the Na⁺:K⁺ ATPase because Na⁺:K⁺ pumps seem to be absent from the outer segment membrane²¹, and because we observe similar currents in the absence of internal ATP or in the presence of 100 μM strophanthidin.

reverse modes of the exchange will be affected equally, leaving the equilibrium level of [Ca²⁺]_i unchanged. After the exchange has reached equilibrium, a tenfold reduction in [K⁺]_o causes a strong activation of inward current, reflecting a movement of Ca²⁺ from the cell, followed by an equal and opposite movement of charge when [K⁺]_o is restored (Fig. 3a, traces for 110 [Na⁺]). We conclude that K⁺ is co-transported by the exchange.

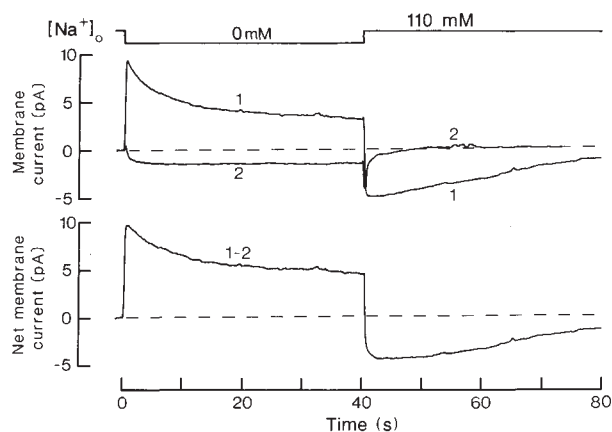
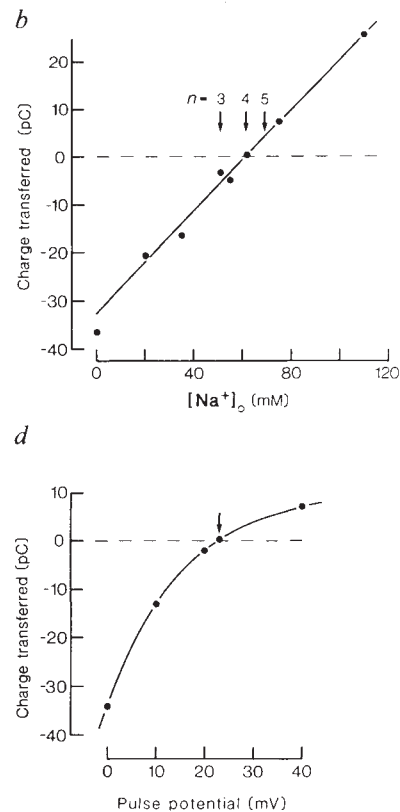
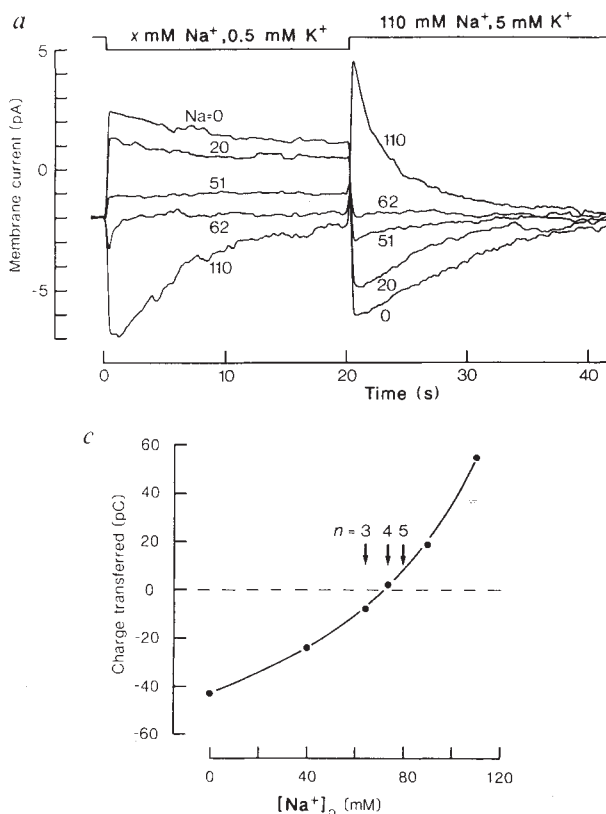


Fig. 2 Effect of external K^+ on reversed exchange currents. *a*, Trace 1 shows the outward current activated by withdrawal of external Na^+ (outer segment maintained in 110 mM Na^+ , 1 mM Ca^{2+} , 5 mM K^+ , 10 mM HEPES, and exposed to 110 mM Li^+ , 1 mM Ca^{2+} , 5 mM K^+ , 10 mM HEPES), and the subsequent inward current flow on restoration of $[Na^+]_o$. When $[K^+]_o$ was also withdrawn (trace 2; rod outer segment exposed to 110 mM Li^+ , 1 mM Ca^{2+} , 10 mM HEPES) a small inward current was seen, but this current was not associated with a substantial Ca^{2+} influx or efflux because little exchange current was observed on restoration of Na^+ . The inward current seen in zero Na^+ , zero K^+ solution probably reflects a small shift in membrane potential occurring because the overall clamped potential, which is held constant, is equal to the sum of membrane potential and any junction potential produced between the flowing $[Na^+] = 0$ solution and the Na^+ -containing solution in the rest of the bath. Similar results were seen in eight experiments. *b*, The difference between traces 1 and 2 in *a*, which gives net exchange current if the junction current was the same in traces 1 and 2. Net K^+ -sensitive charge efflux was 215 pC, and the influx on restoration of Na^+ was 190 pC. In comparison, the charge influx on restoration of Na^+ after exposure to zero $[Na^+]_o$ in the absence of external K^+ (trace 2 in *a*) was 6 pC.

Fig. 3 *a*, Currents observed during a 20 s exposure to a solution containing 0.5 mM K^+ , 1 mM Ca^{2+} and the stated concentration of Na^+ ; the rod outer segment was otherwise maintained in 110 mM Na^+ , 5 mM K^+ and 1 mM Ca^{2+} . Note that the time constant of relaxation increases in low $[Na^+]_o$, so that the relaxation in (for instance) 20 mM $[Na^+]_o$ is slower than that on return to 110 mM $[Na^+]_o$. The brief inward current on exposure to 62 mM $[Na^+]_o$ is an artefact caused by cell movement during the solution change. This artefact is less prominent both in the other traces shown in Fig. 3*a* and in similar experiments on other rods. *b*, Relation between $[Na^+]_o$ during the exposures shown in *a* and charge flow on return to 110 mM $[Na^+]_o$. The straight line, drawn by eye, crosses the axis at $[Na^+]_o = 62$ mM. The arrows show the predicted crossing points for exchange stoichiometries of $n Na^+ : 1 K^+$, where $n = 3, 4$ and 5. Similar results were obtained in four experiments, with a mean $[Na^+]_o$ for zero charge transfer of 62.7 ± 0.8 mM (mean $\pm$ s.e.m.). *c*, A similar experiment in which $[Ca^{2+}]_o$ and $[Na^+]_o$ were changed simultaneously. The rod outer segment was maintained in 110 mM Na^+ , 5 mM Ca^{2+} , 5 mM K^+ ; $[Ca^{2+}]_o$ was reduced to 1 mM and $[Na^+]_o$ reduced to values shown on the abscissa for 100 s. The ordinate shows charge transferred on return to control solution. The continuous curve, drawn by eye, crosses the axis at $[Na^+]_o = 72$ mM. Arrows show the expected null points for $n Na^+ : 1 Ca^{2+}$, where $n = 3, 4$ and 5. Similar results were obtained in three experiments, with the mean $[Na^+]_o$ for zero charge transfer of 73.0 ± 0.6 mM. *d*, The effect of a simultaneous change in membrane potential and $[K^+]_o$ on charge transfer. The rod outer segment was maintained in 110 mM Na^+ , 1 mM Ca^{2+} , 5 mM K^+ and $[K^+]_o$ reduced to 2 mM for 40 s. Depolarizing pulse coincided with the solution change. The ordinate shows the charge transferred on return to control conditions and the abscissa the pulse amplitude (from a holding potential of -14 mV). The arrow shows the expected change in V_m to null a 2.5-fold change in $[K^+]_o$ if one charge exchanges for one K^+ ion. Pulse lengths in these experiments were not long enough to allow equilibrium to be reached in all conditions, and the experimental points would not therefore be expected to fit equation 1 except at the null point. The deviations are particularly noticeable for low $[Na^+]_o$ or for large depolarizations, when the time constant of equilibration is long.



The other traces in Fig. 3a show the effects of simultaneous reductions in $[Na^+]_o$ and $[K^+]_o$. If n Na^+ ions exchange for one K^+ ion then the exchange will remain at equilibrium if the ratio $[Na^+]^n/[K^+]$ remains constant. The solutions with $[Na^+]_o = 51$ mM and 62 mM were chosen to be near equilibrium for a tenfold change in $[K^+]_o$ and for $n = 3$ and 4 respectively, and it can be seen from Fig. 3a that the equilibrium is maintained at $[Na^+]_o = 62$ mM, whereas $[Na^+]_o = 51$ mM causes a charge efflux, with a corresponding influx on return to $[Na^+]_o = 110$ mM (plotted as a function of $[Na^+]_o$ in Fig. 3b). The intersection

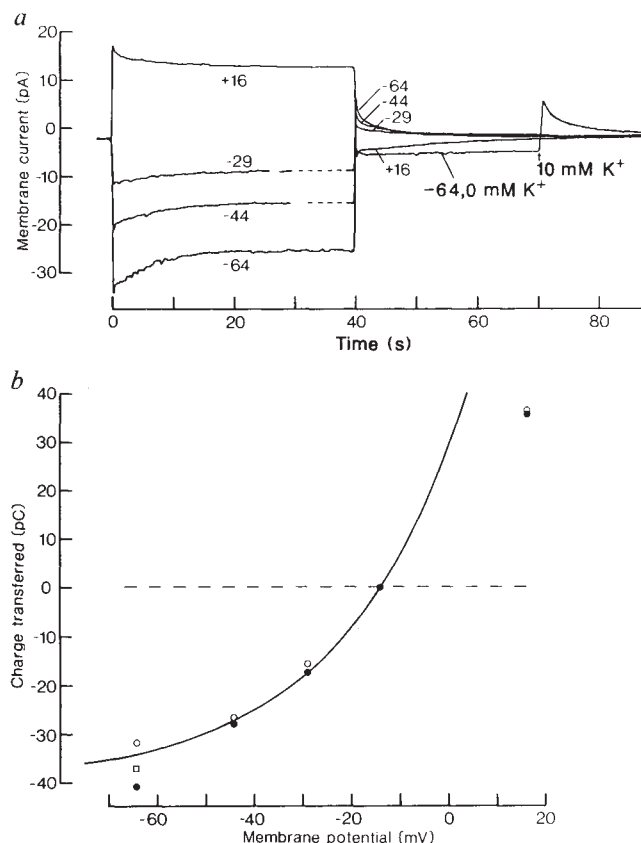


Fig. 4 *a*, Effect of membrane potential on exchange currents. V_m was displaced from resting level of -14 mV to the levels shown. Pulses to -29 mV and -44 mV lasted for 30 s; repolarizations were displaced on the time axis, as indicated by the dotted line, for comparison with other traces. The rod outer segment was maintained in 110 mM Na^+ , 1 mM Ca^{2+} , 10 mM K^+ throughout, except for the trace labelled -64 mV, 0 K^+ when it was stepped to 110 mM Na^+ , 1 mM Ca^{2+} , 0 mM K^+ at the onset of hyperpolarization, and was returned to 110 mM Na^+ , 1 mM Ca^{2+} , 10 mM K^+ as shown by the arrow. *b*, The relation between pulse potential and charge transferred by the exchange during the pulse (filled circles) or on return to $V_m = -14$ mV (open circles; values multiplied by -1). Note that, as in Fig. 1, the charge moved during the pulse is equal and opposite to the charge moved on return to control conditions. The open square shows charge transfer on readmission of 10 mM K^+ (arrow in *a*), which is similar to that moved on repolarization from -64 mV in the presence of K^+ , showing that (1) reversed exchange depends on the presence of external K^+ (see Fig. 2); and (2) the principal route for Ca^{2+} entry is via the exchange, as there is no significant Ca^{2+} entry when the exchange is inactivated in zero $[K^+]_o$. In a separate series of experiments, the time constant of loss of a known calcium load was found to be more than 100 s when the exchange was inactive. The continuous curve is an exponential function rising e -fold in 25 mV, which provides a reasonable fit to the experimental points. The deviation of the point at $+16$ mV from the curve arises at least in part because the time constant with which the exchange reaches equilibrium is longer at positive potentials, and the duration of the pulse was insufficient to allow full equilibration.

with the abscissa in Fig. 3b occurs near the value of $[Na^+]_o = 61.9$ mM expected for a 4 Na^+ :1 K^+ exchange.

Figure 3c shows a similar equilibrium experiment in which $[Ca^{2+}]_o$ and $[Na^+]_o$ were altered simultaneously. A fivefold decrease in $[Ca^{2+}]_o$ was balanced by a decrease in $[Na^+]_o$ from 110 mM to 72 mM, close to the value of 73.6 mM expected for a 4 Na^+ :1 Ca^{2+} exchange. In Fig. 3d $[K^+]_o$ was reduced by a factor of 2.5 and the membrane potential was simultaneously depolarized. The 2.5-fold reduction in $[K^+]_o$ was exactly balanced by a 23 mV depolarization, as expected if one net charge is exchanged with every K^+ .

These experiments show that 4 Na^+ exchange for one K^+ (Fig. 3a, b), that 4 Na^+ exchange for one Ca^{2+} (Fig. 3c) and that one K^+ exchanges for one positive charge (Fig. 3d). The only stoichiometry consistent with each of these measurements is 4 Na^+ :1 Ca^{2+} , 1 K^+ , where the colon indicates countertransport and the comma co-transport. This is consistent with the charge moved after a change in $[Ca^{2+}]_o$, whereas a stoichiometry of 3 Na^+ :1 Ca^{2+} overestimates the charge flow (see Fig. 1 legend). It is also consistent with the known countertransport of one positive charge for one Ca^{2+} (refs 4–7). We note, however, that our results are not obviously compatible with the finding that in cardiac sarcolemmal vesicles 3 Na^+ ions are co-transported with one positive charge⁸.

In the absence of other energy sources, the equilibrium level of such an electrogenic exchange process must depend on membrane potential, with an e -fold change in free $[Ca^{2+}]_i$ occurring for every 25 mV change in membrane potential. Figure 4a shows that hyperpolarization activates an inward exchange current and that depolarization has the reverse effect. The displacement of charge as a function of pulse potential is shown in Fig. 4b, and the experimental observations of charge displacement both during and after the pulse are consistent with an e -fold increase in intracellular calcium per 25 mV depolarization.

A stoichiometry of 4 Na^+ :1 Ca^{2+} , 1 K^+ should enable the exchange to attain a much lower level of $[Ca^{2+}]_i$ than previously supposed. If Na^+ and K^+ do not compete for each other's binding sites then the equilibrium level of $[Ca^{2+}]_i$ is given^{2,9} by:

$$[Ca^{2+}]_i = [Ca^{2+}]_o \frac{[Na^+]_i^4 [K^+]_o}{[Na^+]_o^4 [K^+]_i} \exp(V_m F / RT) \quad (1)$$

In the rod outer segment, in bright light, the membrane potential, V_m , is about -55 mV (ref. 10), and taking $[Na^+]_i = 10$ mM¹¹ and $[K^+]_i = 110$ mM, we obtain an equilibrium level in Ringer's solution of $[Ca^{2+}]_i = 1.8 \times 10^{-10}$ M, nearly 500 times lower than the value of 8.5×10^{-8} M for a 3 Na^+ :1 Ca^{2+} exchange. Although it seems unlikely that this equilibrium level would be attained in practice, an exchange stoichiometry in which K^+ is transported enables the exchange to continue extruding Ca^{2+} to produce a much lower $[Ca^{2+}]_i$ than was previously supposed. In the rod outer segment, where $[Ca^{2+}]_i$ inhibits the guanylate cyclase at $[Ca^{2+}]_i \approx 10^{-7}$ M^{12,13}, the ability to attain a low $[Ca^{2+}]_i$ may be particularly important in the control of the light-sensitive current¹⁴.

In other cells there is also evidence for an interaction of K^+ ions with the exchange. In squid axon¹, bovine rod outer segments¹⁵ and ventricular myocytes¹⁶ the reversed exchange depends on an alkali metal ion in the external solution, with K^+ being considerably more effective than Li^+ (ref. 17). The forward mode of the exchange in squid axon¹⁸ and bovine rod outer segments¹⁹ depends on internal K^+ . We conclude that an exchange of Na^+ for Ca^{2+} and K^+ may be a general phenomenon, although in the rod outer segment the K^+ -binding site may have evolved a higher degree of selectivity to maintain Ca^{2+} extrusion at low $[Ca^{2+}]_i$ without the assistance of an ATP-dependent Ca^{2+} pump.

We thank Professors A. L. Hodgkin and I. M. Glynn for their comments on the manuscript. This work was supported by the MRC, EMBO and NATO. P.McN. is a Nuffield Foundation Science Research Fellow.

Received 14 December 1988; accepted 16 January 1989.

1. Baker, P. F. Blaustein, M. P., Hodgkin, A. L. & Steinhardt, R. A. *J. Physiol.* **200**, 431-458 (1969).
2. Blaustein, M. P. & Hodgkin, A. L. *J. Physiol.* **200**, 497-527 (1969).
3. Reuter, H. & Seitz, N. *J. Physiol.* **195**, 451-470 (1968).
4. Yau, K.-W. & Nakatani, K. *Nature* **311**, 661-663 (1984).
5. McNaughton, P. A., Cervetto, L. & Nunn, B. J. *Nature* **322**, 261-263 (1986).
6. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol.* **391**, 347-370 (1987).
7. Lagnado, L., Cervetto, L. & McNaughton, P. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4548-4552 (1988).
8. Reeves, J. P. & Hale, C. C. *J. biol. Chem.* **259**, 7733-7739 (1984).
9. Hodgkin, A. L. & Nunn, B. J. *J. Physiol.* **391**, 371-398 (1987).
10. Baylor, D. A., Matthews, G. & Nunn, B. J. *J. Physiol.* **354**, 203-224 (1984).
11. Torre, V. *J. Physiol.* **333**, 315-341 (1982).
12. Lolley, R. N. & Racz, E. *Vision Res.* **22**, 1481-1486 (1982).
13. Koch, K.-W. & Stryer, L. *Nature* **334**, 64-66 (1988).
14. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol.* **358**, 447-468 (1985).
15. Schnetkamp, P. P. M. *J. Physiol.* **373**, 25-45 (1986).
16. Gadsby, D. C., Nakao, M., Noda, M. & Shepherd, R. N. *J. Physiol.* **407**, 135P (1988).
17. Allen, T. J. A. & Baker, P. F. *J. Physiol.* **378**, 53-76 (1987).
18. DiPolo, R. & Rojas, H. *Biochim. biophys. Acta* **776**, 313-316 (1984).
19. Schnetkamp, P. P. M., Szerencsei, R. T. & Basu, D. K. *Biophys. J.* **53**, 389a (1988).
20. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **395**, 6-18 (1981).
21. Spencer, M., Detwiler, P. B. & Bunt-Milam, A. H. *Invest. Ophthalmol. Vis. Sci.* **29**, 1012-1020 (1988).

Direct binding of influenza peptides to class I HLA molecules

Benjamin P. Chen & Peter Parham

Department of Cell Biology, Stanford University, Stanford, California 94305, USA

Activation of T lymphocytes requires the intracellular fragmentation of foreign antigens and their presentation by class I or class II major histocompatibility complex (MHC) glycoproteins¹⁻². The direct binding of peptides to class II molecules has been demonstrated using equilibrium dialysis, gel filtration and fluorescence energy transfer at planar membranes, and its specificity compared to that of T-cell activation³⁻⁶. In contrast, direct binding of peptides to class I molecules has been difficult to detect; although peptide sensitization experiments⁷⁻⁸ and the crystallographic structure of HLA-A2 (ref. 9) persuasively argue for its occurrence and importance. Here we describe a gel filtration assay from which we derive direct evidence for selective binding of an influenza matrix peptide to HLA-A2 and for binding of an influenza nucleoprotein peptide to HLA-B37. These two peptides have previously been shown to act respectively as targets for certain HLA-A2 or HLA-B37 restricted influenza-specific cytotoxic T lymphocytes (CTL)^{7,8,10}. In addition we demonstrate binding to some, but not all, HLA allospecificities that cannot present these peptides to CTL. We estimate that less than 0.3% of the HLA molecules present in any given purified preparation were able to bind the added peptides.

A synthetic peptide corresponding to residues 56-68 of the influenza matrix protein (FLU-M1) was radioiodinated and aliquots were incubated with affinity-purified preparations of detergent solubilized HLA-A2, HLA-Aw69 and HLA-B5. Each mixture was then applied to a Sephadex G50 column and the eluted fractions assayed for peptide by direct counting of radioactivity (Fig. 1b) and for class I HLA molecules by indirect radioimmunoassay (Fig. 1c). A small peak of radioactive peptide was found in the excluded volume of all samples, including the negative control with peptide alone. A slightly increased amount of excluded radioactivity was consistently recovered from samples containing HLA-A2 and HLA-Aw69 compared to HLA-B5. The effect was small, however, and raised questions as to whether the excluded radioactivity actually represented peptide bound to HLA molecules. The FLU-M1 peptide is hydrophobic (Fig. 1a), relatively insoluble in aqueous solutions and the excluded peaks of radioactivity could represent self-aggregates or the partitioning of peptide into detergent micelles. Either of these two interpretations would account for the small radioactive

peak seen in the excluded volume of the peptide control (Fig. 1b).

To distinguish between specific binding to class I MHC molecules and such non-specific effects, class I HLA molecules were purified from fractions corresponding to the excluded peak of the G50 column using an anti-class I HLA monoclonal antibody affinity column. Radioactive peptide which was bound and eluted from such columns could only derive from a direct interaction of peptide with a class I HLA molecule. This analysis revealed the presence of FLU-M1/HLA complexes, although the radioactivity recovered was in every case only a small fraction (0.1-1.3%) of that present in the excluded peak of the G50 column (Fig. 2a). Although losses in the extensive manipulation are likely, these results indicate that much of the radioactivity in the excluded peak was not specifically bound to class I molecules and that G50 chromatography alone is not a reliable indicator of complex formation.

After this second step of purification selectivity in the binding of FLU-M1 to different HLA-A, B molecules was observed (Fig. 2a). Two to threefold more radioactivity was consistently recovered in the HLA-A2 and HLA-Aw69 preparations compared to the HLA-B5 preparation which was only slightly more radioactive than the control performed in the absence of HLA (Fig. 2a). Similar binding specificity with HLA-A2, -Aw69 and -B5 was observed in two replicate experiments. A fourth experiment gave no binding to either HLA-A2, -Aw69 or -B5, but this was probably a consequence of the particular peptide preparation used as this was freshly iodinated for each experiment. The selective binding observed in these experiments is consistent with the reported ability of this peptide to sensitize uninfected HLA-A2 cells to killing by influenza-specific CTL, and the more recent evidence for the capacity of HLA-Aw69 to present FLU-M1 to specific CTL (Dr Andrew McMichael personal communication and ref. 11).

Similar results were obtained on replacing the immunoaffinity column with antibody coated onto magnetic beads, and we used this modified, simpler, procedure to compare the binding of the FLU-M1 peptide and FLU-NP, an influenza nucleoprotein peptide (residues 335-349), to various HLA preparations. Previous analysis of CTL target specificity showed that FLU-NP is presented by HLA-B37 (refs 8 and 10) in contrast to FLU-M1 which is presented by HLA-A2 and -Aw69 (refs 7, 10, 11). Three experiments gave comparable results and one is shown in Table 1.

The FLU-M1 peptide bound to HLA-A2 and -Aw69 but not to HLA-B5 and -B44; results similar to those obtained with the affinity column. In addition FLU-M1 showed strong binding to a mixed preparation of HLA-A1 and -B37 molecules. This binding can be primarily attributed to HLA-B37 as similar results were obtained with magnetic beads coated with W6/32, a monoclonal antibody that binds HLA-A, B and C, and with Tü48, a monoclonal antibody with polymorphic specificity that should only bind to HLA-B37 molecules in the mixed preparation¹³.

The FLU-NP peptide showed a different pattern of binding thereby providing evidence for the specificity of the interactions. Greatest binding was obtained with HLA-B37, the molecule known to present FLU-NP to CTL, but significant binding was also seen with HLA-B44 and to a lesser extent with HLA-A2. The binding of both FLU-M1 and FLU-NP to HLA-A2 is unlikely to be due to a greater degree of non-specific binding to HLA-A2 molecules compared to HLA-B5 and -B44 as sixteen other peptides of different origins and unknown immunogenicity showed no binding to HLA-A2 (data not shown). The same sixteen peptides were not tested for binding to HLA-B37.

For both peptides we observed strong binding to the appropriate restriction molecules, HLA-A2 and -Aw69 for FLU-M1, HLA-B37 for FLU-NP. In addition we have demonstrated binding of FLU-M1 to HLA-B37 and FLU-NP to HLA-B44 and HLA-A2, molecules that, as far as is known, do not present these peptides to CTL. Such binding without presentation has

Fig. 1 Separation of free FLU-M1 peptide from HLA and HLA-FLU-M1 complexes by gel filtration. *a*, Amino-acid sequence of the peptide corresponding to residues 56–68 of the influenza matrix protein; a tyrosine residue has been added at the amino terminal for radioiodination and a cysteine residue was added at the carboxyl terminal to enable conjugation. *b*, Radioactivity contained in 1 ml G50 column fractions of HLA and peptide mixtures. *c*, *d*, Fractions from the G50 column assayed for class I HLA molecules using a solid-phase sandwich radioimmunoassay (*c*, Peptide + HLA-A2 column; *d*, Peptide alone column). The open bars give control binding in the absence of the second monoclonal anti-HLA-A, B, C antibody, the closed bars give binding in its presence. The profiles for the column with peptide + HLA-Aw69 and peptide + HLA-B5 were similar to that with peptide + HLA-A2.

Methods. *a*, The FLU-M1 peptide was synthesized by a solid phase method using Fmoc chemistry¹⁴, purified by reverse-phase high performance liquid chromatography, sequenced by automatic Edman degradation using a gas phase sequencer and shown to be >90% pure. The peptide was iodinated by the chloramine T method and in each experiment a separate freshly iodinated preparation of peptide was used. Calculations based on the amount of peptide and the incorporation of radioactivity indicate that 100% of the peptide molecules were iodinated. *b*, HLA molecules were purified from detergent-solubilized cell lysates by immunoaffinity chromatography as previously described¹⁵. HLA-A2 and Aw69 molecules were purified from cell lysates of human lymphoblastoid B cell lines 721.53 (A2) and IDF (Aw69, A26; B18, Bw38) respectively using a monoclonal anti-A2, Aw69 (CR11.351) (ref. 16) antibody column. HLA-B5 molecules were purified from cell lysates of B-lymphoblastoid cell line 721.144 (B5) using a monomorphic anti-HLA A, B, C (W6/32) (ref. 17) monoclonal antibody column. HLA-A2, Aw69 or B5 (0.2 μ M) was taken up in 0.5% deoxycholic acid in Tris-buffered saline (DOC-TBS) containing 1 mM phenylmethylsulphonylfluoride, 0.26 mg ml⁻¹ *o*-phenanthroline, 50 μ g ml⁻¹ pepstatin A, 50 mM iodoacetamide and 3 mg ml⁻¹ edetic acid. Aliquots of ¹²⁵I-labelled FLU-M1 peptide (0.1 μ M) were added to each HLA preparation and the mixtures, and a control of peptide in buffer alone, were incubated for 60 h at room temperature to allow for complex formation. The HLA-peptide complexes were separated from free peptide by gel filtration on a Sephadex G50 column. The column was eluted in 0.5% DOC-TBS, 1 ml fractions were collected and radioactivity contained in each fraction was measured. *c*, The sandwich radioimmunoassay was performed in microtitre plates pre-coated for 24 h at 4 °C with F(ab')₂ fragments of a monomorphic, monoclonal anti-HLA-A, B, C antibody MB40.5 (γ 1) (0.2 μ g per well) (refs 18, 19). The wells were then filled with 1% bovine serum albumin in phosphate-buffered saline (BSA-PBS) and incubated for 1 h at 4 °C. Aliquots (100 μ l) of Sephadex G50 column fractions were added to duplicate wells and incubated for 24 h at 4 °C. After four washes with BSA-PBS, 20 μ l of BB7.7, a second monomorphic anti-HLA-A, B, C monoclonal IgG_{ab} antibody that is directed against a topographically distinct epitope from MB40.5 (ref. 18), was added and the plates were incubated for 1 h at 4 °C. The wells were washed as before and 3 $\times$ 10⁵ c.p.m. of ¹²⁵I-labelled, affinity purified goat anti-mouse IgG₂ (¹²⁵I-GAM₂) added. After further incubation for 1 h at 4 °C, the wells were washed again and assayed for radioactivity. The ¹²⁵I-GAM₂ shows no binding to F(ab')₂ fragments of MB40.5. As positive and negative controls, BB7.7 and ¹²⁵I-GAM₂ were added to MB40.5 F(ab')₂ pre-coated wells incubated with 0.2 μ g HLA-A2 and buffer respectively.

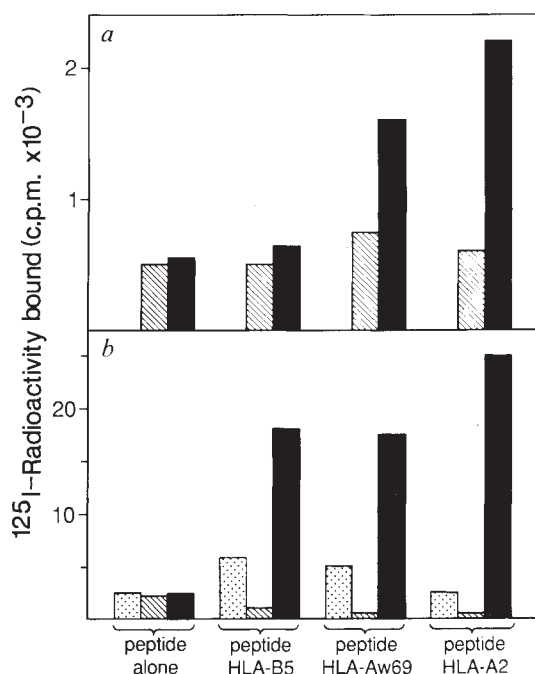
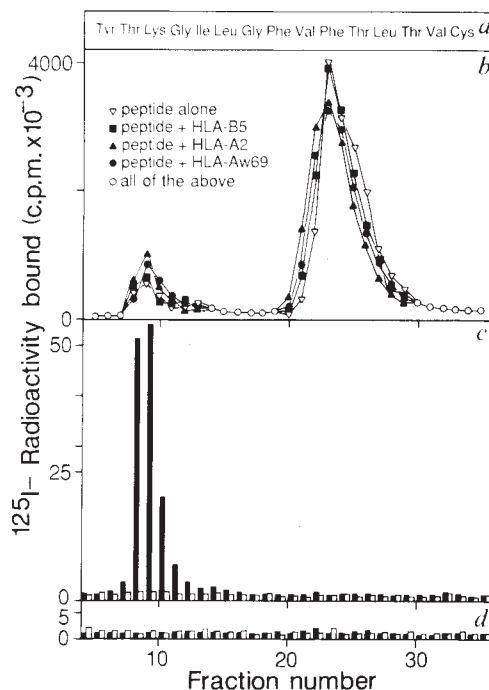


Fig. 2 Isolation of HLA-peptide complexes by anti-class I HLA MAb affinity column. *a*, Radioactivity contained in 3 ml fractions of pre-elution wash (hatched) and an equal volume eluted from the column by elution buffer (dark toned) are shown. *b*, Aliquots (200 μ l) of the unbound fractions (dotted), pre-elution wash (hatched) and fractions eluted from affinity column (dark-toned) were assayed for HLA molecules by the indirect radioimmunoassay described in Fig. 1.

Methods. G50 fractions of the excluded peaks containing HLA molecules were pooled and loaded onto an anti-class I HLA MAb W6/32 affinity column. The column was then washed extensively with 5 column volumes of 0.5% DOC-TBS buffer, followed by an equal volume of 0.5% octylglucoside-TBS buffer to reduce the radioactivity detectable in the wash volumes to approaching background levels. Column-bound HLA molecules and HLA/FLU-M1 complexes were then eluted by an elution buffer (0.1 M NaHCO₃, 0.5M NaCl, 1% octylglucoside, pH 11.3) into tubes containing 0.1 M Tris-HCl. Three ml fractions of wash and eluted samples were collected and measured for radioactivity.

Table 1 Binding of FLU-M1 and FLU-NP peptides to HLA-A, B molecules

		Radioactivity per sample fraction in counts per minute											
Peptide	Fraction	A2		Aw69		B5		B44		A1, B37		Peptide only	
		W6/32	Tü 48	W6/32	Tü 48	W6/32	Tü 48	W6/32	Tü 48	W6/32	Tü 48	W6/32	Tü 48
FLU-M1	Wash	590	620	623	450	883	304	405	680	715	603	646	432
	Elute	7,101	552	6,480	601	904	449	764	455	10,162	9,608	724	340
	Elute-Wash	6,511	0	5,857	151	21	145	359	0	9,447	9,005	78	0
FLU-NP	Wash	3,173	1,411	3,401	2,113	3,690	2,138	3,604	2,208	2,533	1,199	3,230	2,502
	Elute	6,371	2,669	4,561	1,989	3,763	3,265	8,400	10,754	9,938	11,620	4,153	2,819
	Elute-Wash	3,198	1,258	1,160	0	73	1,127	4,796	8,546	7,405	10,421	924	317

The amino-acid sequence of the peptide FLU-M1 is shown in Fig. 1. The sequence for the peptide corresponding to residues 335–349 of the influenza nucleoprotein is Tyr-Ser-Ala-Ala-Phe-Glu-Asp-Leu-Arg-Val-Leu-Ser-Phe-Ile-Arg-Gly. A tyrosine residue was added at the amino terminus to permit radioiodination. HLA-A2, Aw69 and B5 molecules were isolated by affinity chromatography from the cell lines described in Fig. 1. HLA-B44 molecules were isolated from a class I null CIR line transfected with the B44 gene. HLA-A1 and B37 molecules were isolated from a homozygous EBV-LCL line KAS-011 (A1, B37). The separation of free peptide from HLA and HLA-peptide complexes was performed as described in Fig. 1. The HLA-peptide complexes were isolated by indirect immunoprecipitation using W6/32 or Tü 48 monoclonal antibodies together with goat-anti-mouse Ab coated on Dynal magnetic beads (Robbins Scientific, California). W6/32 Ab binds to all the HLA molecules used whereas Tü 48 binds to the Bw4 epitope present on HLA-B5, -B44 and B37 but not A2, Aw69 or A1 (ref. 13). Immunomagnetic beads with HLA-peptide complexes bound were washed extensively with 0.5% DOC-TBS buffer and complexes were eluted from the beads with 0.1 M NaHCO₃, 0.5 M NaCl, 1% octylglucoside, pH 11.3.

frequently been seen in studies on the binding of peptides to murine class II MHC molecules⁵. The binding of one peptide to two or more class I molecules is presumably a reflection of shared properties of their peptide binding grooves; the lack of presentation could be due either to holes in the T-cell receptor repertoire or to inefficient binding of the naturally processed antigen fragments.

From the amounts of radioactivity and of HLA-A or B recovered from the affinity columns or immunomagnetic beads, we estimate that no more than 1 in 300 HLA-A2, -Aw69, -B37 and -B44 molecules were binding peptide in these experiments and the actual fraction could be an order of magnitude less. A much higher proportion of class II MHC molecules can be shown to bind an appropriate peptide, although it is never a majority of molecules^{3,4}. Kinetic analysis shows this is not due to a simple mass-action effect as the complexes once formed are extremely stable¹². The alternative interpretation is that the extent of binding is limited by the presence of previously bound peptide, an idea that is supported by the presence of extra density in the combining site of crystallized HLA-A2 molecules⁹. The results presented here are consistent with this hypothesis and would also suggest that dissociation of endogenous peptides from class I molecules is generally slower than for class II.

These experiments demonstrate selective binding of two antigenic peptides to class I MHC molecules. The effects are small and their detection required a double selection procedure. The approach used has the advantage of generating purified peptide/MHC complexes which are then available for further study and should allow characterization of the binding site specificity of class I HLA molecules and of the interaction of peptide/class I complexes with the T cell receptor. Such studies would however, be greatly facilitated by methods to increase the fraction of class I molecules binding a defined peptide.

This work is supported by grants from the NIH. P.P. is a scholar of the Leukemia Society of America; B.P.C. is a fellow of the Cancer Research Institute. We thank K. Drickamer for peptide sequence analysis.

Received 28 December 1988; accepted 19 January 1989.

1. Townsend, A. R. M. *Immunol. Rev.* **6**, 80–100 (1987).
2. Unanue, E. R., Beller, D. I., Lu, C. Y. & Allen, P. M. *J. Immun.* **132**, 1–5 (1984).
3. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359–361 (1985).
4. Buus, S., Sette, A., Colon, S., Miles, C. & Grey, H. M. *Science* **235**, 1353–1358 (1987).
5. Guillet, J. G., Lai, M. Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* **324**, 260–262 (1986).
6. Watts, T. H., Gaub, H. E. & McConnell, H. M. *Nature* **320**, 179–181 (1986).
7. Gotch, F., Rothbard, J., Howland, K., Townsend, A. & McMichael, A. *Nature* **326**, 881–882 (1987).
8. Townsend, A. R. M., et al. *Cell* **44**, 959–968 (1986).
9. Bjorkman, P. J. et al. *Nature* **329**, 506–512 (1987).
10. Gotch, F., McMichael, A., Smith, G. & Moss, B. *J. exp. Med.* **165**, 408–416 (1987).

11. Bodmer, H. C., Gotch, F. M. & McMichael, A. J. *Nature* **337**, 653–655 (1989).
12. Buss, S., Sette, A., Colon, S., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071–1077 (1986).
13. Müller, C., Ziegler, A., Müller, G., Schunter, F. & Wernet, P. *Human Immun.* **5**, 269–281 (1982).
14. Sheppard, R. C. & Williams, B. J. *Int. J. Peptide Protein Res.* **20**, 451–454 (1982).
15. Parham, P. *Meth. Enzym.* **92**, 110–138 (1982).
16. Russo, C., Ng, A. K., Pellegrino, M. A. & Ferrone, S. *Immunogenetics* **18**, 23–35 (1983).
17. Barnstable, C. J. et al. *Cell* **14**, 9–20 (1978).
18. Brodsky, F. M. & Parham, P. *J. Immun.* **128**, 129–135 (1982).
19. Parham, P. *J. Immun.* **131**, 2895–2902 (1983).

Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis

Ann M. Stock, James M. Mottonen, Jeffry B. Stock & Clarence E. Schutt

Departments of Chemistry and Biology, Princeton University, Princeton, New Jersey 08544, USA

Homologies among bacterial signal transduction proteins suggest that a common mechanism mediates processes such as chemotaxis, osmoregulation, sporulation, virulence, and responses to nitrogen, phosphorous and oxygen deprivation^{1–3}. A common kinase-mediated phosphotransfer reaction has recently been identified in chemotaxis, nitrogen regulation, and osmoregulation^{4–10}. In chemotaxis, the CheA kinase passes a phosphoryl group to the cytoplasmic protein CheY, which functions as a phosphorylation-activated switch that interacts with flagellar components to regulate motility. We report here the X-ray crystal structure of the *Salmonella typhimurium* CheY protein. The determination of the structure was facilitated by the use of site-specific mutagenesis to engineer heavy-atom binding sites. CheY is a single-domain protein composed of a doubly wound five-stranded parallel β -sheet. The phosphoacceptor site in CheY is probably a cluster of aspartic-acid side chains near the C-terminal edge of the β -sheet. The pattern of sequence similarity of CheY with components of other regulatory systems can be interpreted in the light of the CheY structure and supports the view that this family of proteins have a common structural motif and active site.

Crystals of native *S. typhimurium* CheY protein were obtained using vapour diffusion techniques with ammonium sulphate as precipitant. The wild-type protein gave clusters of well-formed, elongated prisms at 1.8–2.0 M ammonium sulphate, pH 4.5. The space group of the crystals was orthorhombic, $P2_12_12_1$, with one CheY monomer per asymmetric unit, and unit cell

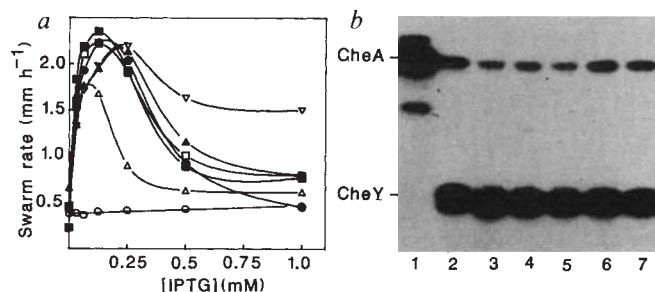


Fig. 1 CheY proteins containing cysteine/serine substitutions are active *in vivo* and *in vitro*. *a*, *S. typhimurium* PS257 (*cheY*62 *srl-202::Tn10recA1*) containing wild-type and cysteine-substituted *cheY* genes inserted behind the regulatable *lac* promoter moter in pMLB1120.215-derived plasmids were assayed for chemotaxis. Cells were grown and inoculated into the centre of semi-solid agar plates (0.3% agar, 1.3% tryptone, 9.7% NaCl, 40 μ g ml⁻¹ ampicillin). The plates were incubated at 37 °C and rates were determined from four swarm diameter measurements taken over a 10-hour period. Expression of *cheY* was induced to varying levels by including the indicated concentrations of isopropyl β -D-thiogalactopyranoside (IPTG) in the tryptone agar. Plasmid inserts: no insert, (○); wild-type CheY, (●); S15C, (△); S56C, (▲); S76C, (□); S79C, (■); S104C, (▽). *b*, Pure [³²P]phospho-CheA, (200 nM), was incubated with pure CheY proteins (200 nM) for 30 s at 20 °C as described previously⁴. The reaction (20 μ l) was quenched by addition of SDS sample buffer, and applied to a 15% polyacrylamide-SDS gel. An autoradiograph of the electrophoretogram is shown. Lane, no CheY; 2, wild-type CheY; 3, S15C; 4, S56C; 5, S76C; 6, S79C; 7, S104C.

Methods. A 0.8 kilobase (kb) *Sna*BI-*Sma*I fragment containing the *S. typhimurium cheY* gene was inserted into the polylinker region of M13mp10 (ref. 27) and used as a template for oligonucleotide-directed mutagenesis^{28,29}. Cysteine codons were substituted for serine codons at sites corresponding to serines 15, 56, 76, 79 and 104, and the mutations confirmed by nucleotide sequencing³⁰. The 0.8 kb fragments containing the altered *cheY* genes were moved into two vectors: a plasmid containing a regulatable *lac* promoter, pMLB1120.215 (obtained from M. Berman, Litton Institute of Applied Biotechnology, Rockville, Maryland), and a high copy plasmid, pUC12 (ref. 27). CheY proteins S15C, S56C, S76C, S79C and the wild-type protein, were purified using a modification of previously described methods¹ from soluble extracts of *E. coli* HB101 cells³¹ containing the pUC12-derived plasmids. CheY S104C partitioned entirely into inclusion bodies and soluble, active protein was obtained by renaturation of the aggregated protein. CheY S104C was recovered from a sonicated cell extract by centrifugation at 10,000 g, solubilized in 6.0 M guanidine hydrochloride, 5.0 mM Tris-Cl, 0.1 mM dithiothreitol, pH 8.0, and dialysed overnight at 4 °C versus 5.0 mM Tris-Cl 0.1 mM dithiothreitol, pH 8.0. The soluble fraction was applied to a DE-52 column, and the concentrated protein eluted with 0.5 M NaCl in Tris/dithiothreitol buffer and further purified by gel filtration chromatography.

parameters $a = 45.9$ Å, $b = 47.2$ Å and $c = 54.2$ Å. There has been a report of a monoclinic crystal form of the *Escherichia coli* CheY protein¹¹. Despite a difference of only three amino acids between the sequences of the *E. coli* and *S. typhimurium* CheY proteins^{1,12}, the conditions used to crystallize the former, Tris-HCl pH 8.0 as buffer, and ammonium sulphate as precipitant, produced only masses of uncharacterizable needles with the *S. typhimurium* protein.

A major stumbling block in protein crystallography is often the trial-and-error search for suitable heavy-atom derivatives, a process potentially facilitated by modern genetic techniques¹³. There are no cysteine residues in CheY, and it was therefore possible to engineer mutant proteins with single cysteine substitutions as potential sites for mercury binding. Using oligonucleo-

Table 1 Crystallographic data

Space group: $P2_12_12_1$ ($a = 45.9$ Å, $b = 47.2$ Å, $c = 54.2$ Å)			
Data set S56C	Native	Ethylmercury chloride	Uranyl acetate
No. of observations	17,871	14,580	6,735
No. of unique observations	3,167	3,257	2,688
No. $>3\sigma$ (%)	95.5	92.6	93.7
Resolution limit (Å)	2.7	2.7	2.7
R_{merge} (%)†	7.9	7.5	7.1
Mean fractional isomorphous change (%)		18.6	21.1
$k_{\text{emp}}^{\ddagger}$		7.1	4.7
Phasing power§		1.67	1.81
Correlation coefficient¶		0.43	0.51
$R_{\text{centric}}^{\P}$		53.1	53.3
Heavy-atom sites (fractional coordinates)		site 1	site 1 site 2
X		0.9014	0.0934 0.6269
Y		0.8777	0.5623 0.1017
Z		0.7595	0.8418 0.7872
Occupancy (arbitrary units)		8.82	13.00 5.05
Temperature factor (Å ²)		20.0	32.5 43.0
Mean figure of merit: 0.60			

Crystals of CheY were obtained by the vapour diffusion method using protein at a concentration of approximately 10 mg ml⁻¹ in 50.0 mM sodium acetate, pH 6.0. The reservoir solution containing 1.8 M ammonium sulphate, 1% polyethylene glycol 400, and 50 mM sodium acetate, pH 4.5 was mixed with an equal volume of protein solution, generally 10 μ l each, and then left undisturbed for 3–4 days at 4 °C. All solutions used with the cysteine-substituted mutants also contained 0.1 mM dithiothreitol. Wild-type CheY produced clusters of prisms from which single crystals of approximately 0.1 $\times$ 0.1 $\times$ 1.0 mm could be isolated. The mutant S56C produced prisms that were consistently thicker than wild-type, typically 0.2 $\times$ 0.2 mm. All data in this study were obtained using the Enraf-Nonius FAST system (Fast Scanning Area Sensitive TV Detector Diffractometer). The X-ray source was an Elliot GX-21 rotating anode generator operating at 35 kV, 35 mA with a graphite monochromator. Crystals were rotated about the c^* axis by more than 90°, collecting 0.1° frames of 60 or 90 s each. The crystal-to-film distance was normally 50 mm with a detector tilt angle of 10°, which allows data to be collected to about 2.4 Å resolution. Data evaluation and collection used the program MADNES²⁴. Calculation of unit cell constants and preliminary screening of heavy-atom derivatives were both also done entirely with the FAST system. All crystallographic calculations were performed using the CCP computing package kindly provided by Dr Phil Evans (MRC, Cambridge). Heavy-atom parameters were refined against the centric data using the F_{HLE} method²⁵.

† R_{merge} = conventional discrepancy R-factor for scaling symmetry-related intensities, in this case from single crystals.

‡ k_{emp} is twice the ratio of isomorphous to anomalous mean intensity differences²⁶.

§ Phasing power = root mean square f_{H}/E .

¶ Correlation coefficient is between f_{H} and observed ΔF (ref. 25).

¶ R_{centric} = conventional discrepancy R-factor for comparing observed and calculated isomorphous differences of the centric terms.

tide-directed mutagenesis of the cloned *cheY* gene, we introduced cysteine residues at each of the positions occupied by serine in the wild-type protein. Five mutant proteins were obtained, designated S15C, S56C, S76C, S79C and S104C. To test the physiological activity of the altered proteins, the mutated genes, inserted behind regulatable promoters on multicopy plasmids, were introduced into a null *cheY* mutant. All the cysteine-containing proteins were similar to the wild-type protein in their ability to restore chemotactic behavior to the CheY-deficient strain (Fig. 1a). All of the purified cysteine-substituted CheY proteins exhibited phosphotransfer activity in assays with pure phospho-CheA (Fig. 1b).

Despite the apparently normal activities of the cysteine-containing CheY proteins, the single-atom sulphur for oxygen

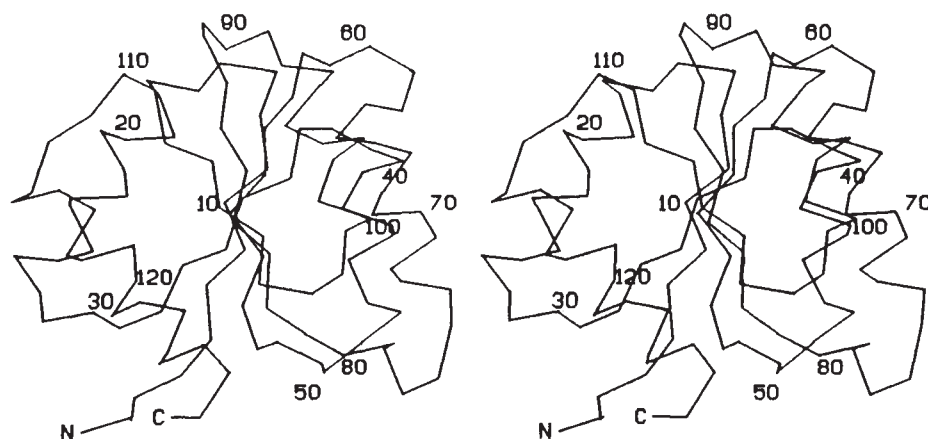


Fig. 2 Stereo image of the CheY C α backbone. The structure consists of a five-stranded parallel β -sheet flanked on both sides by α -helices. The N and C termini, corresponding to residues 2 and 129, are indicated by N and C; the chain is numbered at every tenth residue.

substitutions did have some effects. Three of the mutant proteins (S15C, S76C and S104C) aggregated into inclusion bodies to varying extents, and the soluble forms of these proteins tended to precipitate rather than crystallize¹⁴. Another mutant (S79C) was soluble but, unlike the wild-type, crystallized in the trigonal space group $P3_121$ or $P3_221$, with unit cell parameters $a = b = 61.5$ Å and $c = 85.1$ Å. The S79C crystals were much larger than the wild-type prisms and may prove useful in future high-resolution studies. Only one of the five cysteine/serine substitutions (S56C) produced crystals similar to the crystals of wild-type CheY. The space group and unit cell parameters for crystals of this mutant protein were virtually identical to those of the wild-type protein. Moreover, the R_{scale} (see Table 1) between wild-type and mutant data sets at 2.7 Å resolution was 0.091, a value in the same range as the R_{merge} calculated from any single crystal. These data, together with the demonstration of physiological function and phospho-acceptor activity (Fig. 1) establish that the structure of S56C is equivalent to that of the wild-type protein. Crystals of the mutant protein were used in collecting both native and derivative diffraction data.

The cysteine-substituted CheY protein, S56C, reacted readily with ethylmercury chloride, yielding a single-site derivative. A second derivative, obtained by reaction of the crystals with uranyl acetate, was solved by difference Fouriers using combined isomorphous and anomalous phases from the mercury derivative. X-ray diffraction data collected from crystals of the S56C mutant, and ethylmercury chloride and uranyl acetate derivatives are summarized in Table 1. An electron density map calculated using anomalous and isomorphous information from both derivatives showed clear regions of α -helix and β -sheet. The initial chain tracing was aided by localization of the engineered cysteine at the mercury-binding site. The major uranyl site was close to the carboxylate groups of Asp 12 and Asp 13, and a minor uranyl-binding site was near Asp 41. Although the chain was easily followed, the protruding region formed by residues 90–109 was generally of weaker density, possibly indicating inherent disorder or mobility in this part of the molecule. The R -factor for the current model is 0.29 at 2.7 Å resolution.

CheY is a single-domain protein with a doubly wound five-stranded α/β motif¹⁵. A central parallel β -sheet flanked on both sides by α -helices is formed from five β -strands ($\beta 1$ – $\beta 5$) and five α -helices (αA – αE) alternating along the primary sequence (Fig. 2). The topology of the β -sheet is $\beta 2, \beta 1, \beta 3, \beta 4, \beta 5$. Looking down the sheet, which has the usual right-handed twist, the helices are arranged at roughly even spacings along the outside of the molecule, with αA and αE packing against one face of the sheet, αB and αC against the other face, and αD sitting alongside $\beta 5$ and protruding somewhat from the rest of the molecule.

CheY is homologous over its entire length to the N-terminal domains of more than fifteen different bacterial proteins that regulate adaptive responses to changing environmental condi-

tions⁶. Sequence similarities between these domains can be interpreted in terms of the CheY structure. Sequence alignments reveal three stretches of four hydrophobic residues at sites corresponding to the three internal β -strands of CheY (Fig. 3a). Another significant conserved feature is the occurrence of hydrophobic amino acids at positions that correspond in CheY to the internal faces of helices that pack against the β -sheet. This conserved hydrophobic core within the family of CheY homologues indicates that there is a common structural motif.

CheY and its homologues seem to function as phosphorylation-activated response regulators that control a wide range of different cellular activities. The mechanism established in chemotaxis (Che)^{6,7} and nitrogen regulation (Ntr)⁹ involves transfer of a phosphoryl moiety from imidazole nitrogens in histidine protein kinases (CheA and NtrB) to carboxylate oxygens in response regulators (CheY and NtrC). The similarity of these reactions is emphasised by the 'crosstalk' between components involved in different signal transduction pathways (phosphotransfer from CheA to NtrC and from NtrB to CheY)¹⁶. Since CheY and its homologues share a common chemical mechanism, they should also share a common active site.

Although comparisons between any two proteins in the CheY family show 20–30% sequence identity, only six amino-acid residues are completely conserved among all the members of the set. The conserved residues Gly102 and Ala103 are found in a turn at the beginning of $\beta 5$; Asp13 at the end of $\beta 1$, Asp57 at the end of $\beta 3$ together with Lys109 and Pro110 at the end of $\beta 5$, are all located on the same face of the molecule at the C-terminal ends of the strands of the parallel β -sheet (Fig. 3b). Asp13, Asp57 and Asp12 (a highly conserved residue which is either Asp or Glu in other members of the family) are clustered in an acidic pocket formed by loops at the C-terminal ends of $\beta 1$ and $\beta 3$ (Fig. 3c). The location of these acidic residues suggests an active site, by analogy with other doubly-wound parallel β -sheet domains in which binding sites are almost invariably formed by loops connecting the C-terminal ends of adjacent β -strands to helices lying on opposite sides of the sheet¹⁷. As activation of CheY involves the transfer of a phosphoryl moiety from a histidine side chain of CheA to carboxylate side chains of CheY, it seems likely that the acidic cluster functions as the site of phosphorylation in the CheY protein.

The structure and function of the chemotaxis response regulator may have general implications for the role of acyl phosphates in regulatory biochemistry. Phosphotransfer reactions between protein phosphoramidates and protein acyl phosphates have so far been observed only in the signal transduction pathways of bacterial regulatory systems. But high-energy phosphorylation of protein carboxylate side chains, although rare compared to the lower-energy phosphorylation of side-chain hydroxyl groups, has been observed. The best characterized examples are the aspartyl phosphates that are central to the mechanism of ATP-driven ion-transport systems such as the

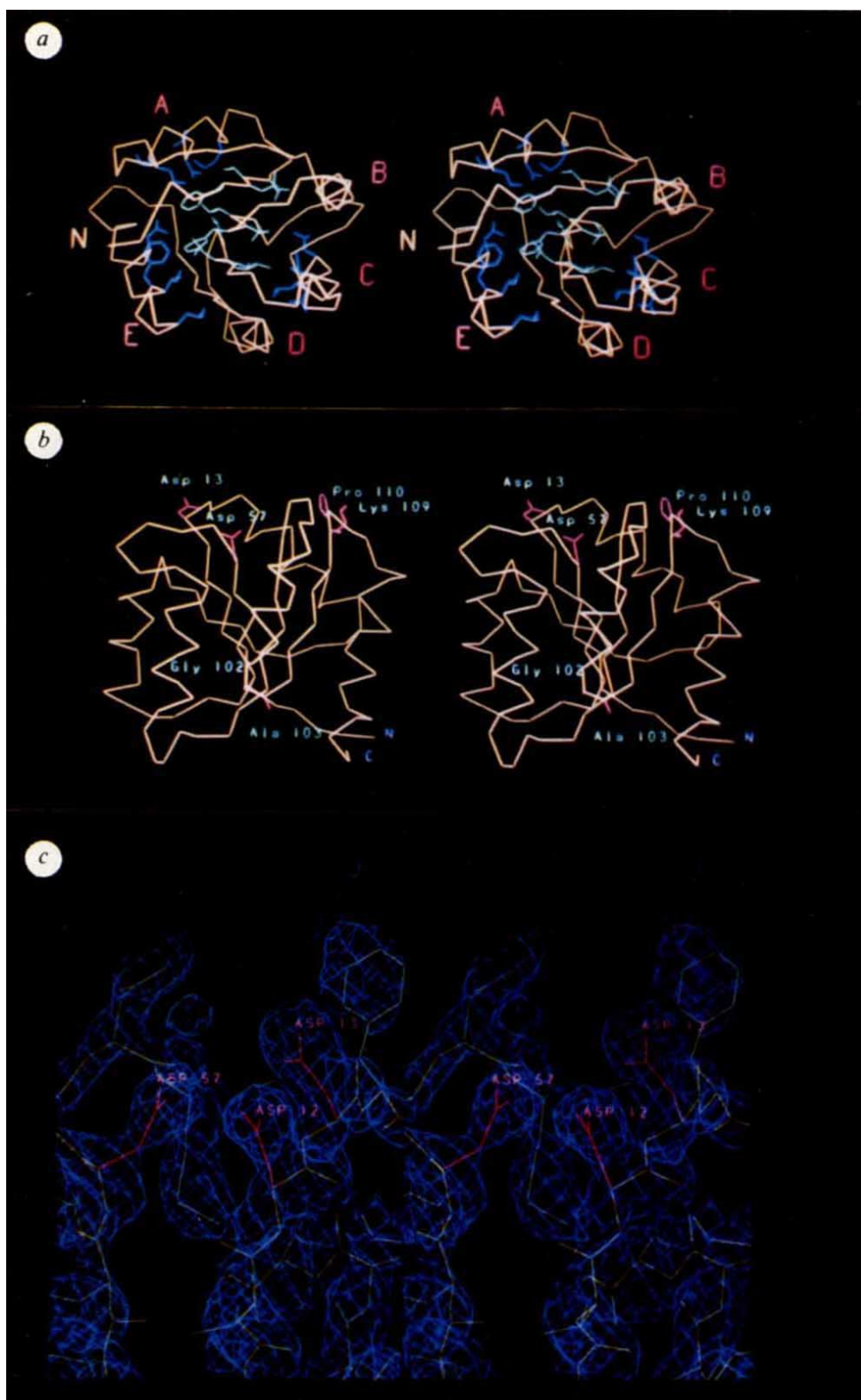


Fig. 3 *a*, Stereo image of CheY showing hydrophobic side chains that are conserved in the family of bacterial regulatory proteins homologous to CheY. The C α skeleton is shown in gold. Side chains contributed by the three central parallel β -strands, β 1 (Phe8, Leu9, Val10, Val11), β 3 (Phe53, Ile54, Ile55, Cys56), and β 4 (Val83, Leu84, Met85, Val86), are shown in green; those contributed by helices α A (Met17, Ile20, Val21, Leu24, Leu25), α C (Leu66, Leu68, Leu69, Ile72) and α E (Leu120, Ile 123, Phe124, Leu127) are shown in blue. This view looks down from the top of the molecule as shown in Fig. 2. Letters A-E identify the five α -helices and N indicates the N-terminus. *b*, Stereo image of CheY showing positions of the six residues (Asp 13, Asp 57, Gly 102, Ala 103, Lys 109, Pro 110) that are identical in the CheY family of bacterial regulatory proteins. The conserved side chains are displayed in pink. Note that the two conserved aspartate residues appear at the C-terminal ends of adjacent parallel β -strands that fold into helices on opposite sides of the sheet as is common for active sites in other parallel α/β proteins. *c*, Stereo image showing the $2F_o - F_c$ electron density map in the region near the putative phospho-acceptor site. The three aspartate residues (Asp 12, Asp 13, Asp 57) that constitute the acidic pocket are shown in red.

Na $^{+}$, Ca $^{2+}$, H $^{+}$ and K $^{+}$ pumps that control the ionic composition of the cytosol, and are largely responsible for transmembrane electrochemical gradients¹⁸⁻²¹. A structure for the Ca $^{2+}$ ATPase has been proposed from primary sequence analyses¹⁹ in which the phosphorylated aspartate residue is located in an extramembrane domain that has an α/β motif similar to that of CheY. Phosphorylation of aspartate side chains in the ATPases leads to conformational changes that control ion movements across membranes^{22,23}. It seems likely that an analogous phosphorylation-induced conformational change functions in CheY and its

homologues to regulate motility and control gene expression in bacterial cells. The CheY structure may therefore represent a phosphorylation-activated switch that provides a general mechanism for stimulus-response coupling in biological systems.

We thank Gregory Petsko for his support and encouragement, and James Myslik and Katrina Forest for technical assistance. J. M. M. is a Princeton Visiting Scholar from MIT. A.M.S. is a Fellow of the Damon Runyon-Walter Winchell Cancer Research Fund. This work was supported by grants from the NIH (to J.B.S. and C.E.S.), American Heart Association (to C.E.S.) and

the Trustees of Princeton University (to C.E.S.). A.M.S. and J.M.M. contributed equally to the work in this manuscript.

Received 6 September 1988; accepted 6 January 1989.

1. Stock, A. M., Koshland, D. E., Jr. & Stock, J. B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7989-7993 (1985).
2. Nixon, B. T., Ronson, C. W. & Ausubel, F. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7850-7854 (1986).
3. Kofoed, E. C. & Parkinson, J. S. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4981-4985 (1988).
4. Wylie, D. C., Stock, A., Wong, C.-Y. & Stock, J. *Biochem. biophys. Res. Commun.* **151**, 891-896 (1988).
5. Hess, J. F., Oosawa, K., Kaplan, N. & Simon, M. I. *Cell* **53**, 79-87 (1988).
6. Stock, A. M. *et al. Cold Spring Harb. Symp. quant. Biol.* **53**, (in the press).
7. Hess, J. F., Bourret, R. B. & Simon, M. I. *Nature* **336**, 139-143 (1988).
8. Ninfa, A. J. & Magasanik, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5909-5913 (1986).
9. Weiss, V. & Magasanik, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8919-8923 (1988).
10. Igo, M. M. & Silhavy, T. J. *J. Bact.* **170**, 1571-1573 (1988).
11. Volz, K., Beman, J. & Matsumura, P. *J. biol. Chem.* **261**, 4723-4725 (1986).
12. Matsumura, P., Rydel, J. J., Linzmeier, R. & Vacante, D. J. *Bact.* **160**, 36-41 (1984).
13. Dao-Pin, S., Alber, T., Bell, J. A., Weaver, L. H. & Matthews, B. W. *Prot. Eng.* **1**, 115-123 (1987).
14. Krueger, J. K., Stock, A. M., Schutt, C. E. & Stock, J. B. in *AAAS Symposium on the Protein Folding Problem* (in the press).
15. Richardson, J. S. *Meth. Enzym.* **115**, 341-358 (1985).
16. Ninfa, A. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5492-5496 (1988).
17. Brändén, C.-I. *Q. Rev. Biophys.* **13**, 317-338 (1980).
18. Skou, J. C., Norby, J. G., Maunsbach, A. B. & Esmann, M. (eds.) *Prog. in Clin. and Biol. Res.* **268A**, (1987).
19. Brandl, C. J., Green, N. M., Korczak, B. & MacLennan, D. H. *Cell* **44**, 597-607 (1986).
20. Serrano, R., Kielland-Brandt, M. C. & Fink, G. R. *Nature* **319**, 689-693 (1986).
21. Hesse, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4746-4750 (1984).
22. Jencks, W. P. *Adv. Enzymol.* **51**, 75-106 (1980).
23. Tanford, C. *CRC Crit. Rev. Biochem.* **17**, 123-152 (1984).
24. Messerschmidt, A. & Pflugrath, J. *Appl. Crystallogr.* **20**, 306-315 (1987).
25. Dodson, E. J. in *Crystallographic Computing Techniques* (ed. Ahmed, F.R.) 259-268 (Munksgaard, Copenhagen, 1976).
26. Dodson, E. & Vijayan, M. *Acta crystallogr.* **B27**, 2402-2411 (1971).
27. Messing, J. *Meth. Enzym.* **101**, 20-78 (1983).
28. Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-500 (1983).
29. Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).
30. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
31. Bolivar, F. & Backman, K. *Meth. Enzym.* **68**, 245-267 (1979).

A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence

Gustavo A. Gonzalez, Karen K. Yamamoto, Wolfgang H. Fischer, David Karr, Patricia Menzel, William Biggs III, Wylie W. Vale & Marc R. Montminy

The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

Cyclic AMP regulates the expression of a number of genes through a conserved promoter element, the CRE¹. Moreover, transcriptional induction by cAMP requires the activation of cAMP-dependent protein kinase (protein kinase A)^{1,2}. We have previously characterized the cAMP response element binding protein (CREB) in PC12 cells and brain tissue as a nuclear factor, of relative molecular mass 43,000, whose transcriptional efficacy is regulated by protein kinase A phosphorylation^{3,4}. CREB stimulates transcription on binding to the CRE as a dimer. Experiments suggesting that the dimerization and transcriptional efficacy of CREB are each stimulated by phosphorylation at distinct sites prompted us to suggest that CREB is regulated by multiple kinases *in vivo*⁴. We now report the isolation of a cDNA clone for rat CREB using amino-acid sequence information from purified CREB protein. Sequence analysis of this CREB cDNA predicts a cluster of protein kinase A, protein kinase C and casein kinase II consensus recognition sites near the N terminus of the protein. The proximity of these potential phosphorylation sites to one another indicates that they may interact either positively or negatively to regulate CREB bioactivity.

About 100 pmol (5 µg) CREB protein was purified from rat brain tissue as previously described⁴. After preparative SDS-

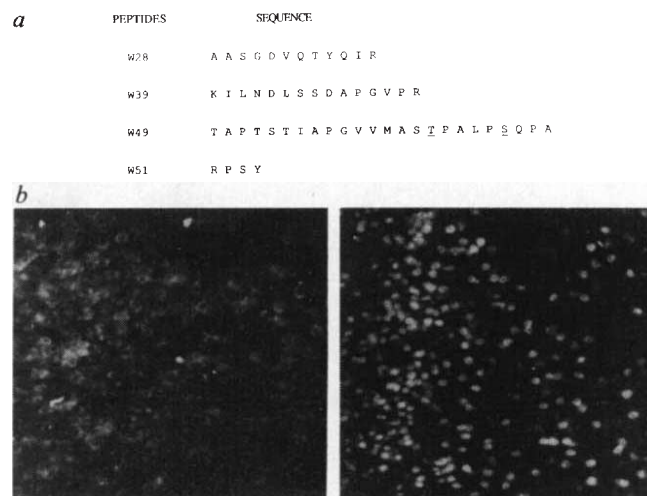


Fig. 1 *a*, Amino-acid sequence of CREB fragments. Four peptides, W28, W39, W49 and W51 are shown. Misassigned residues, underlined, were tentatively identified at levels below 0.5 pmol. *b*, Indirect immunofluorescence microscopy of thin slices from pyriform cortex treated with W39 antiserum. Left, micrograph of cells incubated with preabsorbed W39 antiserum. Right, micrograph of cells incubated with W39 antiserum.

Methods. Nitrocellulose-bound CREB was digested with trypsin (Worthington) and fragments were separated by reversed-phase HPLC (Vydac C-18 column, 4.6 × 250 mm, pore size: 300 Å, particle size: 5 µm) using a Hewlett-Packard 1084 B Liquid Chromatograph at a flow rate of 1 ml min⁻¹ (mobile phase: 0.05% trifluoroacetic acid in water-acetonitrile with increasing concentration of acetonitrile). Fractions were collected manually based on absorbance at 210 nm. Fragments were analysed by Edman degradation on an Applied Biosystems Protein Sequencer 470A equipped with an online PTH Analyzer 120A from the same manufacturer. Between 15 and 0.5 pmol PTH amino acid were detected per cycle. The identity of phosphoserine in peptide W(51) was deduced from the presence of the PTH dehydroalanine-dithiothreitol adduct in position 3 and the radioactivity of the fragment. In *b*, W39 peptide was synthesized using the primary sequence data in Fig. 1*a*. This peptide was conjugated to human α-globulins and antisera were prepared as described¹⁷. Cryostat sections (20 µm) of rat brain were incubated in a 1:1,500 dilution of primary W39 antiserum. The primary antibody was visualized after reaction with fluorescein isothiocyanate-conjugated anti-rabbit immunoglobulin¹⁸. Preabsorbed antiserum was prepared by incubating 100 µg of W39 peptide per 50 µl antiserum overnight at 4°C. Before use, the immunoglobulin conjugate was also preabsorbed.

PAGE this protein was transferred to nitrocellulose and visualized with amido black. The 43K band was excised and digested with trypsin⁵. Three peptide fragments, called W28, W39 and W49, were identified by HPLC and sequenced (Fig. 1*a*). A fourth tryptic fragment, peptide W51, containing the protein kinase A phosphorylation site⁴, was isolated after labelling CREB with protein kinase A and [³²P]ATP. The amino-acid sequence of the radiolabelled peptide is consistent with that of a protein kinase A phosphorylation site (Arg-Arg-X-Ser)⁶.

To confirm that these tryptic peptides were fragments of the CREB protein, we developed antisera against the synthetic W39 peptide and used indirect immunofluorescence to determine the subcellular distribution of the antigen recognized by W39 antiserum. Using cryostat sections of pyriform cortex from rat brain, we observed a nuclear staining pattern which was blocked by prior treatment of the antiserum with synthetic W39 peptide (Fig. 1*b*). To determine the relative molecular mass of this antigen, we prepared nuclear extracts from PC12 cells that had been labelled *in situ* with inorganic ³²P. After immunoprecipitation and SDS-PAGE analysis, we observed a single radiolabelled immunoreactive product of 43K (Fig. 2*a*). In contrast, this

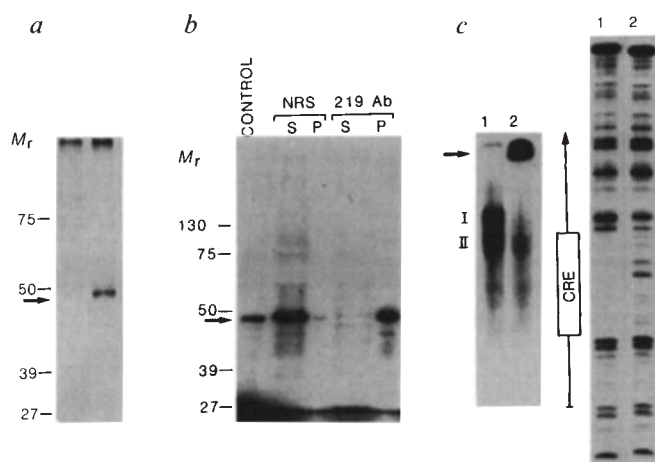


Fig. 2 *a*, Immunoprecipitation of CREB from PC12 cells labelled *in situ* with inorganic ^{32}P . Lane 1, sample precipitated with pre-immune serum. Lane 2, sample precipitated with W39 antiserum. Relative molecular mass in thousands is shown on the left. The arrow points to the 43K phosphoprotein. PC12 cells were labelled and collected as described previously³. Extracts were immunoprecipitated¹⁹ and products were resolved on 10% SDS-polyacrylamide gels. Gels were dried and visualized by autoradiography. *b*, Immunoprecipitation of CREB from partially purified CREB preparations. CREB was labelled with ^{32}P with protein kinase A⁴ and then incubated without serum (control), or with pre-immune or normal rabbit serum (NRS) or W39 antiserum (219 Ab). After immunoprecipitation, samples were divided into supernatant (S) and pellet (P) fractions. Supernatants were further concentrated by precipitating with 10% trichloroacetic acid. After SDS-PAGE, bands were visualized by autoradiography. The arrow points to the 43K phosphoprotein. The relative molecular mass in thousands is shown on the left. *c*, Effect of W39 antiserum on CRE-binding activity. Lane 1, samples treated with preabsorbed W39 antiserum. Lane 2, samples treated with W39 antiserum. Left panel, gel mobility shift assay using partially purified CREB protein and double-stranded ^{32}P -labelled CRE oligonucleotide⁴. Complexes I and II, dimer and monomer forms of CREB, respectively. The arrow points to the high relative molecular mass complex of what is thought to be CREB and anti-W39 bound to CRE oligonucleotide probe. Right panel, DNase I protection assay using the same CREB preparation with ^{32}P -labelled *Ava*II-*Hind*III somatostatin promoter fragment³. Position of the somatostatin CRE are shown.

product was not observed using pre-immune serum. Immunoprecipitation of partially purified CREB preparations labelled *in vitro* with protein kinase A also showed a single 43K band (Fig. 2b), indicating that W39 antiserum specifically recognizes a nuclear phosphoprotein with a relative molecular mass (M_r) identical to that of CREB.

We used gel mobility shift and DNase I protection assays to determine whether W39 antiserum could modify or interfere with CRE-binding activity. Gel mobility shift assays using purified CREB revealed two protein DNA complexes corresponding to monomer and dimer forms of CREB bound to the ^{32}P -labelled CRE-oligonucleotide⁴. Co-incubation with W39 antiserum caused a shift of these complexes to a new complex of higher relative molecular mass (Fig. 2c, left). Formation of this complex was completely blocked by prior treatment of W39 antiserum with synthetic W39 peptide (Fig. 2c, lane 1) or by omission of CREB from the reaction mixture (data not shown). To determine whether W39 antiserum could immunoprecipitate CRE-footprinting activity, we performed DNase I protection assays on supernatants of CREB protein which had been immunoprecipitated with either W39 antiserum, 'blocked' W39 antiserum, or pre-immune serum. Whereas all the CRE-binding activity was retained in supernatants using blocked antiserum

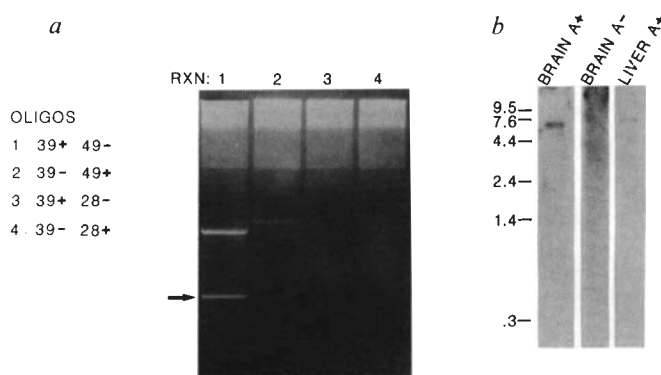


Fig. 3 *a*, Isolation of CREB cDNA fragment by PCR amplification. Synthetic sense (+) and antisense (-) oligonucleotides were prepared for each tryptic fragment (28, 39, 49). Four combinations of sense and antisense oligonucleotides, shown on the left, were used to prime a rat cortex cDNA library (Okayama-Berg pCD1) as template. PCR-extended products for each reaction (1 to 4) were resolved by agarose gel electrophoresis and visualized after staining with ethidium bromide. The arrow points to a 0.4 kb fragment containing the CREB protein sequence. *b*, Northern blot analysis of brain and liver mRNA using ^{32}P -labelled CREB cDNA fragment.

Methods. The following sense-strand oligonucleotides were used for PCR reactions: 28, 5'-GCNGCNCNCGNGAYGTNCAXA-CNTAYCAXAT-3' where N = A + G + T + C, X = A + G, Y = C + T. Primer 39, 5'-AAYGAYCTNA/TG/CNA/TG/CNGAYGCNCC-NGGNGTNC-3'. Primer 49, 5'-ACNATNGCNCNCGNGGTN-GTNATGGC-3'. Antisense oligonucleotides (-) contained the complements of these sequences. The reaction mixtures were in 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 5 mM MgCl_2 , 0.01% (w/v) gelatin, 200 μM each dNTP, 5–10 μg of each oligonucleotide, 2.5 units *Taq* polymerase (Perkin Elmer Cetus), and 2.0 μg rat cortex library cDNA, in a total volume of 100 μl . After overlaying the samples with mineral oil, 35 cycles of denaturation, annealing and extension were completed. Annealing reactions were performed at 60 $^\circ\text{C}$ for 1.5 min, followed by 2.5 min extension at 72 $^\circ\text{C}$, and 1 min denaturation at 94 $^\circ\text{C}$. For the northern analysis, poly(A)⁺RNA (brain A+, liver A+) from brain and liver was prepared and electrophoresed with poly(A)⁺RNA from brain (brain A-). Samples were blotted, incubated with a ^{32}P -labelled 0.4 kb CREB fragment (Fig. 3a) and washed as described previously¹. Hybridizing bands were visualized by autoradiography. Sizes (in kilobases) are shown on the left.

or pre-immune serum, less than 10% of the activity was recovered in supernatants of extracts precipitated with untreated W39 antiserum (Fig. 2c, right). Our results therefore indicate that W39 peptide is indeed a fragment of CREB protein because W39 antiserum specifically recognizes a 43K nuclear phosphoprotein with CRE-binding activity.

Reports showing that the polymerase chain reaction (PCR) procedure could be used to obtain cDNA fragments from amino-acid sequence data^{7,8} (S. Gould, manuscript in preparation), prompted us to construct synthetic 28-base oligonucleotide primers corresponding to the least degenerate regions of each tryptic fragment. Because the relative order of these tryptic fragments within the mature CREB protein was unknown, each possible combination of sense and antisense oligonucleotides from two different fragments was tested using a rat cortex cDNA library as the template. After 35 rounds of amplification, we observed three distinct PCR-extended products by agarose gel electrophoresis (Fig. 3a). DNA sequence analysis revealed that the 0.4-kilobase (kb) fragment contained sequences for an additional five residues of the W49 peptide. Moreover, this fragment also encoded the entire W28 peptide fragment, verifying that it was an authentic CREB cDNA. The 1-kb PCR fragment in lane 1 did not seem to encode any related protein product, however.

Fig. 4 *a*, Nucleotide and predicted amino-acid sequence of a CREB cDNA clone. DNA sequence is numbered on the right. The longest open reading frame, 1,023 nucleotides, begins with a methionine codon at position 40 and ends with termination codon at position 1,063. Predicted amino-acid sequences indicated on right in parentheses. Sequences encoding CREB tryptic peptides W28, W39 and W49 are overlined with arrows. Sequences encoding kinase sites are boxed. The box I encodes protein kinase C site. The box numbered II encodes the W51 peptide and contains phosphorylation motifs for both protein kinase A and C. Box III contains the casein kinase II phosphorylation motif. Leucines thought to form the 'leucine zipper' structure are underlined. *b*, Footprinting activity of *E. coli* lysates containing β -galactosidase-CREB fusion protein. Lane 1, no protein. Lane 2, *E. coli* lysate. *c*, Amino-acid homology of CREB with transcription factor c-jun and the product of the proto-oncogene c-fos²¹. The position of the first amino acid for each protein is shown in parentheses. Vertical lines indicate amino-acid identity and asterisks show conservative substitutions. Leucine residues that are thought to form the zipper structure are in bold.

Methods. Bacteriophage plaques (5×10^5) from a PC12 λ gt11 cDNA library (kindly provided by J. Boulter) were screened as previously described²⁰ with a 0.4 kb CREB cDNA fragment as probe (Fig. 3a). Using the random primer labelling technique, the 0.4 kb fragment was labelled to a specific activity of 10^9 c.p.m. μ g⁻¹. Hybridizing clones were purified and *Eco*R1 inserts were subcloned into pGEM vectors (Promega). CREB cDNA inserts were sequenced on both strands by the chain termination technique. For the footprinting assays the CREB bacterial expression plasmid was constructed by inserting a 2 kb *Stu*I-*Xba*I CREB cDNA fragment into *Sma*I-*Xba*I-digested pUC18. This plasmid encodes a fusion protein that contains the N terminal 10 amino acids of the β -galactosidase gene followed by the C-terminal 210 amino acids of the CREB protein. Footprinting assays were performed as described in Fig. 2.

Northern blot analysis of brain and liver mRNA using the 0.4 kb PCR fragment as a probe revealed a single mRNA species of about 7 kb (Fig. 3b). To obtain a cDNA containing the entire CREB coding region, we screened a λ gt11 PC12 library and obtained six positive plaques from $\sim 5 \times 10^5$ bacteriophage clones. Sequence analysis of one clone of 1.5 kb revealed a single long open reading frame of 1,023 bases encoding 341 amino acids (Fig. 4). The protein predicted by this cDNA contains all four tryptic peptides sequenced from the CREB protein. CREB tryptic peptides W28, W49 and W51 are flanked by arginine or lysine residues, which is consistent with the cleavage specificity of trypsin. The presence of a Gln residue N-terminal to the W28 peptide, however, suggests that chymotrypsin, a contaminant of trypsin preparations, was responsible for this cleavage. The predicted M_r of the CREB protein encoded by this cDNA (37K)

is somewhat less than the M_r of CREB from PC12 cells and brain tissue (43K). This difference may in part be due to the phosphorylation of CREB in these cells.

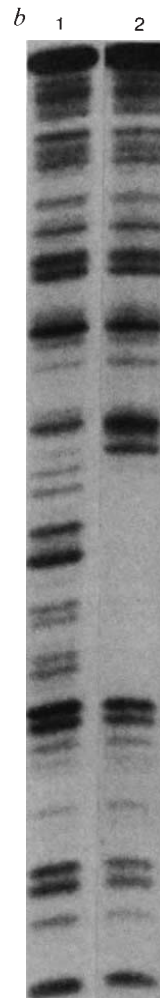
Further inspection of the predicted protein sequence reveals a cluster of potential phosphorylation sites located within the proximal one-third of the CREB molecule. For example, the protein kinase A phosphorylation site which we sequenced within peptide W51 (residues 131-134) is flanked by two arginine residues at its C terminus, thereby creating a protein kinase C phosphorylation motif. This sequence confirms previous *in vitro* phosphorylation experiments with purified CREB which showed that the same protein kinase A site is also phosphorylated by protein kinase C⁴. A second protein kinase C site, also predicted from *in vitro* phosphorylation experiments, is located at residue 121. Nearby we observed the sequence EEEKSEEE starting at

a

GAA AGC AGT GAC TGA GGA GCT TGT ACC ACC GGT GAC TAG ATG ACC ATG GAC TCT GGA GCA	60
Met Thr Met Asp Ser Gly Ala	(7)
GAC AAC CAG CAG AGT GGA GAT GCT GCT GTA ACA GAA GCT GAA AGT CAA CAA ATG ACA GTT	120
Asp Asn Gln Gln Ser Gly Asp Ala Ala Val Thr Glu Ala Glu Ser Gln Gln Met Thr Val	(27)
CAA GCC CAG CCA CAG ATT GCC ACA TTA GCC CAG GTA TCC ATG CCA GCA GCT CAT GCG ACG	180
Gln Ala Gln Pro Gln Ile Ala Thr Leu Ala Gln Val Ser Met Pro Ala Ala His Ala Thr	(47)
TCA TCT GCT CCC ACT GTA ACC TTA GTG CAG CTG CCC AAT GGG CAG ACA GTC CAG GTC CAT	240
Ser Ser Ala Pro Thr Val Thr Leu Val Gln Leu Pro Asn Gly Gln Thr Val Gln Val His	(67)
GGG GTC ATC CAG GCG GCC CAG CCA TCA GTT ATT CAG TCT CCA CAA GTC CAA ACA GTT CAG	300
Gly Val Ile Gln Ala Ala Gln Pro Ser Val Ile Gln Ser Pro Gln Val Gln Thr Val Gln	(87)
TCT TCC TGT AAG GAC TTA AAA AGA CTT TTC TCC GGA ACT CAG ATT TCA ACT ATT GCA GAA	360
Ser Ser Cys Lys Asp Leu Lys Arg Leu Phe Ser Gly Thr Gln Ile Ser Thr Ile Ala Glu	(107)
AGT GAA GAT TCA CAG GAG TCT GTG GAT AGT GTA ACT GAT TCC CAA AAA CGA AGG GAA ATC	420
Ser Glu Asp Ser Gln Glu Ser Val Asp Ser Val Thr Asp Ser Gln Lys Arg Arg Glu Ile	(127)
CTT TCA AGG AGG CCT TCC TAC AGG AAA ATT TTG AAT GAC TTA TCT TCT GAT GCA CCA GGG	480
Leu Ser Arg Arg Pro Ser Tyr Arg Lys Ile Leu Asn Asp Leu Ser Ser Asp Ala Pro Gly	(147)
GTG CCA AGG ATT GAA GAA GAA AAA TCA GAA GAG ACT TCA GCC CCT GCC ATC ACC ACT	540
Val Pro Arg Ile Glu Glu Glu Lys Ser Glu Glu Glu Thr Ser Ala Pro Ala Ile Thr Thr	(167)
GTA ACA GTG CCA ACC CCG ATT TAC CAA ACT AGC AGT GGG CAG TAT ATT GCC ATT ACC CAG	600
Val Thr Val Pro Thr Pro Ile Tyr Gln Thr Ser Ser Gly Gln Tyr Ile Ala Ile Thr Gln	(187)
GGA GGA GCA ATA CAG CTG GCT AAC AAT GGT ACC GAT GGG GTA CAG GGC CTG CAG ACA TTA	660
Gly Gly Ala Ile Gln Leu Ala Asn Asn Gly Thr Asp Gly Val Gln Gly Leu Gln Thr Leu	(207)
ACC ATG ACC AAT GCA GCT GCC ACT CAG CCG GGT ACT ACC ATT CTA CAA TAT GCA CAG ACC	720
Thr Met Thr Asn Ala Ala Thr Gln Pro Gly Thr Thr Ile Leu Gln Tyr Ala Gln Thr	(227)
ACT GAT GGA CAG CAG ATT CTA GTG CCC AGC AAC CAA GTT GTT GTT CAA GCT GCC TCT GGT	780
Thr Asp Gly Gln Gln Ile Leu Val Pro Ser Asn Gln Val Val Gln Ala Ala Ser Gly	(247)
GAT GTA CAA ACA TAC CAG ATT GCG ACA GCA CCC ACT AGC ACC ATT GCC CCT GGA GTT GTT	840
Asp Val Gln Thr Thr Tyr Ile Arg Thr Ala Pro Thr Ser Thr Ile Ala Pro Gly Val Val	(267)
ATG GCG TCC TCC CCA GCA CTT CCT ACA CAG CCT GCT GAA GAA GCA GCA CGA AAG AGA GAG	900
Met Ala Ser Ser Pro Ala Leu Pro Thr Gln Pro Ala Glu Glu Ala Ala Arg Lys Arg Glu	(287)
GTT CGT CTA ATG AAG AAC AGG GAA GCA GCA AGA GAA TGT CGT AGA AAG AAG AAA GAA TAT	960
Val Arg Leu Met Lys Asn Arg Glu Ala Ala Arg Glu Cys Arg Arg Lys Lys Lys Ala Tyr	(307)
GTG AAA TGT TTA GAG AAC AGA GTG GCA GTG CTT GAA AAC CAA AAC AAA ACA TTG ATT GAG	1020
Val Lys Cys Leu Glu Asn Arg Val Ala Val Leu Glu Asn Gln Asn Lys Thr Leu Ile Glu	(327)
GAG CTA AAA GCA CTT AAG GAC CTT TAC TGC CAC AAG TCA GAT TAA TTT GGG ATT TAA ATT	1080
Glu Leu Lys Ala Leu Lys Asp Leu Tyr Cys His Lys Ser Asp	(341)
TTC ACC TAT GTG GTG GAA AAT GGA CTG GAT TGG CCA CAA CCA GAA	1125

c

JUN (263)	K A E R K R M R N R I A A S K C R K R K L E R I A R L E E K V K T L K A Q N S E L A S T A N M L R E
CREB (285)	K R E V R L M K N R E A A R E C R R K K K E Y V K C L E N R V A V L E N Q N K T L I E E L K A L K D
FOS (139)	K R R I R R E R N K M A A A K C R N R R R E L T D T L Q A E T D Q L E D E K S A L Q T E I A N L L K



residue 152 (one-letter amino-acid code), which is consistent with a casein kinase II (CK II) phosphorylation motif⁹. CK II is a ubiquitous enzyme, localized in both the nucleus and the cytoplasm, which is thought to regulate expression of a number of growth factors and proto-oncogenes such as the SV40 large T antigen, adenovirus E1a protein and c-Myc¹⁰.

As CK II has a preference for β -turns near its recognition motif¹¹ the secondary structure of CREB, as predicted by the Chou-Fasman rules¹², may contain β -turns directly N terminal to the serine residue. Indeed, preliminary experiments using purified CK II suggest that CREB is phosphorylated by this enzyme (G.A.G. and M.R.M., unpublished observations). The proximity of this CK II motif to protein kinase A and C sites indicates that they may interact to regulate CREB activity. Moreover, the preference of CK II for acidic residues is in marked contrast to protein kinase A and C sites, which are highly basic. Phosphorylation by CK II may therefore serve to inhibit protein kinase A and C phosphorylation and consequently cAMP-responsive gene transcription.

At a cellular level, CK II activation seems to be associated with serum-stimulated cell growth¹⁰. By contrast, cAMP-dependent protein kinase activation enhances cellular differentiation and loss of proliferation¹³. The cellular phenotype may therefore depend, in part, on alternate phosphorylation of CREB by these kinases. *In vitro* mutagenesis of the cloned CREB cDNA will help to clarify the role of CK II and protein kinases A and C in regulating CREB activity.

To verify that the CREB protein encoded by this cDNA contains DNA-binding activity, we used DNase I protection assays with a truncated fragment of the protein expressed in *Escherichia coli*. To express this protein, we constructed a bacterial expression plasmid that encodes the first 10 amino acids of the β -galactosidase protein, followed by the C-terminal 210 amino acids of the CREB protein. After preparing crude lysates from *E. coli* containing this plasmid, we performed footprinting assays and observed that the fusion protein contained CRE-footprinting activity (Fig. 4b, lane 2) which was indistinguishable from native CREB protein (Fig. 2c). Lysates prepared from *E. coli* containing pUC 18 vector alone did not contain CRE-binding activity (data not shown).

We observed a stretch of basic amino acids characteristic of a DNA-binding domain near the predicted C terminus of the protein. This putative DNA-binding region shares sequence homology with transcription factor c-Jun/AP-1^{14,15} and the proto-oncogene c-fos product (Fig. 4c). The close similarity of CREB and c-Jun/AP1-binding sites to one another had prompted us to speculate earlier that these two factors might be structurally related³. Moreover, the high-affinity binding of the AP-1 protein to the somatostatin CRE (unpublished data) indicated that CREB and AP-1 shared overlapping DNA-sequence specificities.

Alignment of the three protein sequences reveals a discreet block of 50 amino acids with substantial homology. Closer inspection of this region also reveals a conserved series of regularly spaced leucine residues, interspersed with charged amino acids. Such regions can potentially form amphipathic α -helices, also termed leucine zippers¹⁶, which may interact with other proteins through their hydrophobic surfaces. Dimerization of CREB could therefore occur, in part, through monomer-monomer interactions in this region. Alternatively, other factors like c-Fos may bind to CREB at this site and regulate its activity.

We thank S. Gould for advice on the PCR technique, Joan Vaughan for preparation of CREB antiserum, P. Sawchenko and V. Roberts for assistance with immunocytochemistry and J. Rivier for synthesis of peptide 239. We also thank W. Boyle and T. Hunter for help with phosphorylation experiments, and R. Evans for advice and support. This work was supported by the NIH, and conducted in part by the Clayton Foundation for Research, California Division. W.W.V. and M.R.M. are Clayton Foundation Investigators.

Note added in proof: After submission of this manuscript, Hoeffler *et al.*²² reported the structure of a human placental cAMP-responsive DNA binding protein which is homologous to rat CREB.

Received 28 November 1988; accepted 16 January 1989.

1. Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6682-6686 (1986).
2. Riabowol, K. T. *et al. Nature* **336**, 83-86 (1988).
3. Montminy, M. R. & Bilezikian, L. M. *Nature* **328**, 175-178 (1987).
4. Yamamoto, K. K., Gonzalez, G. A., Biggs III, W. H. & Montminy, M. R. *Nature* **334**, 494-498 (1988).
5. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6970-6974 (1987).
6. Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *A. Rev. Biochem.* **56**, 567-613 (1987).
7. Lee, C. C. *et al. Science* **239**, 1288-1290 (1987).
8. Saiki, R. K. *et al. Science* **230**, 1350-1352 (1985).
9. Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. B. *J. biol. Chem.* **262**, 9136-9140 (1987).
10. Figge, J., Webster, T., Smith, T. & Paucha, E. *J. Virol.* **62**, 1814-1818 (1988).
11. Hathaway, G. M. & Traugh, J. A. *Curr. Topics cell. Regul.* **21**, 101-127 (1982).
12. Chou, P. & Fasman, G. D. *Biochemistry* **13**, 222-245 (1974).
13. Trepel, J. B. & Neckers, L. M. *Molec. cell. Biol.* **7**, 2644-2648 (1987).
14. Bohmann, D. *et al. Science* **238**, 1386-1392 (1987).
15. Angel, P. *et al. Nature* **332**, 166-171 (1988).
16. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759-1763 (1988).
17. Vaughan, J. M. *et al. Meth. Enzym.* **168**, 588-617 (1988).
18. Sawchenko, P. E. & Swanson, L. W. *Brain Res.* **210**, 31-51 (1981).
19. Chiu, R., Boyle, W., Meek, J., Smeal, T., Hunter, T. & Karin, M. *Cell* **54**, 541-552 (1988).
20. Montminy, M. R., Goodman, R. H., Horovitch, S. J. & Habener, J. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7161-7165 (1984).
21. Van Straaten, F., Muller, R., Curran, T., van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183-3187 (1983).
22. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242**, 1430-1432 (1988).

A pseudo-exon in the functional human α A-crystallin gene

Cynthia J. Jaworski & Joram Piatigorsky

Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

The frequent correspondence of exons to structural or functional domains in proteins has suggested that many proteins have evolved by modular assembly¹. This idea is supported by examples of apparent exon duplication and by shared domains among both alternatively spliced and completely separate genes¹⁻⁴. During this process it is probable that some combinations of exons would not prove advantageous and would therefore be lost. Here we report that within the active single-copy human gene for α A-crystallin there is a 'pseudo-exon' in the early stages of being extinguished, perhaps the result of a failed experiment in the evolution of this specialized, lens-specific protein.

The protein and gene structures of α A-crystallin, a major structural protein in the ocular lenses of all vertebrates, are highly conserved throughout evolution^{5,6}. The lenses of rodents⁷⁻⁹ and certain other mammals (hedgehog, bat and pika)¹⁰ contain an additional form called α A^{ins}, which is identical to α A except for the insertion of 23 amino acids, the result of alternative RNA splicing of an 'insert exon' (IE) located in the first intron^{11,12}. No evidence for expression of α A^{ins}-crystallin has been found in lens proteins from primates (including humans) or from many other mammalian classes, birds or other vertebrates¹⁰.

We have now found an element in the first intron of the human α A-crystallin gene¹³ closely resembling the alternatively spliced exon of the rodent gene. The DNA sequence of clone p α A-T2, a fragment of the human α A-crystallin gene, is shown in Fig. 1. Comparison with the homologous mouse¹¹ and chicken¹⁴ sequences shows striking similarity in exons 1 and 2. Surprisingly, a third region of similarity between human and mouse (but not chicken) is apparent in the first intron and

Fig. 1 Nucleotide and deduced amino-acid sequence for the genomic fragment in αA -T2. Clone αA -T2 extends from 15 nucleotides upstream of the initiator methionine of the protein-coding region in exon 1 through to 240 nucleotides of the second intron. The DNA sequence of this fragment is presented with the exons displayed against a stippled background and the deduced amino-acid sequence of the exons is shown above the single letter code. Exon junctions are at the same position as in all other species examined, with exon 1 containing amino-acid residues 1 through to 63 and exon 2 corresponding to amino-acid residues 64 through to 104 of the αA polypeptide. The deduced amino-acid sequence of the gene fragment in αA -T2 confirms the assigned order of residues in the tryptic fragments of the human αA -crystallin protein^{19,20}. A region within the first intron corresponding to the alternatively spliced exon of rodent αA -crystallin is in brackets. The relative position of this element is similar in both the human and rodent genes. Underlined codons in the second exon are stop codons that would be brought into frame by an upstream deletion (see text).

Methods. Clone αA -T2 was obtained by subcloning into bacteriophage M13 mp19 a *Kpn*I/*Bam*H1 fragment from λ phage LS77 which was isolated from a human spleen genomic library¹³; LS77 contains the coding region for all three exons of the human αA -crystallin gene. DNA sequencing was performed on both strands by the dideoxy chain-termination method²¹ using oligonucleotide primers made on an Applied Biosystems Model 380B oligonucleotide synthesizer.

M D V T I Q H P W F K R T L G P F Y P S R L F D Q F

1 ACCAAAGCTGAACATGGAGTGGACATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCAGCGGGCTGTTCGACCACTT

F G E G L F E Y D L L P F L S S T I S P Y Y R Q S L F R T V

91 TTTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCCCTTCTCTGCTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGCGACCGT

L D S G I S E

181 GCTGGACTCCGGCATCTCTGAGGTAAAGACGTGCCCGTGGTGGCTCTCTCGCTGCTCAGAGGTGGTGGCTCGGTGGTGGAGAG

271 CGATGGACTCTGGTCTTGGCTCCGTGAGGAGGTGGCTCGTCCACTTCATCCCTTGCAGAGGTGGGCGAGAGCCTGTGTCCCACTG

[L

361 CAGCCAGCTGGCAGAGCTTCCCTGGCACTGGGAGAGGGTGGACAAGGAGGAGCCTGAATCCACCTTTCCTTCTCCATCAGCTCA

M T H V C F V R H O P H T G N P K S S P S R]

451 TGACCCATGTGTGCTTGTGAAGGACACGACATCTGGAACCCCAAGAGCAGCCATCCAGGCATCGGTGGTGGCAATGCCAGCTCC

541 CGGGTTCTCTGGTCTCTGAGTCCCGGAGACCTGGGAGCAGGTGGGGTCTATGCTCTGAAAGCCAGAGAGCAGGGCGTTCTAGCACCC

631 TCTCCAATGAGCTCGGCTCGCCACGGTTAGCAAGCTCTTGGCAAGTTTACTTAGGTGCCCTGCAAGGCTAAAGGGACAGGCAATGG

721 ACGCCCCCCCCCAACCAACACAGGCTCTCTCTGAGCCACGGGTGAGCGGTGACAGTTCTGCTGTCTGGAGGGCTGAGTCCAC

811 CCAGCACCTCATAAACAGGGTCTCCCGAGGGCTGCTGAGTACGATCAACGCCAGGGTGCATAATGCCTCAGGAGGCCAAGGCTGAG

901 CCAGGGGAGTGAGAAGGAGCATGTGGAAGTGGCTTTTGGAGAGGAGCTGCGCAGGCTGTGAGCAGGCTCCGGCGCTTCTATAGACAGC

991 ATGACACCAAGGCGAGTACCTCATTCCACAGGCTGAGTCCAGGCCAGGCCAGCCAGCATCAGCAGCAGACGATTGACCTAACGGACC

1081 AACCAACCCGTAACGACCCCTCTACCATACCAAGTACGAGCAGGCCATACCAAGCAGCACTTATCTATAACGAGCCACCTGACCATAG

1171 CCAACAACCCAGCGGGCCCAAGTACGATTACGCCCCCTGAGGTTTGGAGACAGGTCGAGGGTCTGCTGTCTGTCTG

1261 AGGAGACAGTACAGGGCCCCCAAGGCTGCCCCACTTGGTGTGTGGGAGAAGAGGCCGCGAGGTACCGAAGCATCTGTCTGTGATA

V R S

1351 ACCGGGACCCGCTCTCTGCAACCCAGCAGGAGCGGACCTCTGGGAGCTCCACATGGCAGCTTGGATTTCAGGTTCCATCCG

D R D K F V I F L D V K H F S P E D L T V K V Q D D F V E I

1441 ACCGGGACAAGTTCGTATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCAGCGTGAAGGTCAGGAGGACTTGTGGAGATCC

H G K H N E R Q

1531 ACGGAAGGACACAGGAGCGCCAGGTGAGCCAGGCACTGAGAGTGGGACAGGGGGGAGTGGGCGGAGSACAAGGGGTACGGGG

1621 GGCACGACCGGGCTGCACACCTGCACCAATGCCTTCAACCTGGGAGAGGGAACGCTCTCCAGGGGACCCGAAATCAGGCGTGGCTTT

1711 CCGCAAGGAGGGGCGCTGCCACCTGAGCACAGCCAGCCCTCCCGTGACAGAGGTACCATTCGCGAGCTAATGTGGCTCAGGGATC

a	
human	CTCATGACCCATGTGTGTTTGTAAAGGCACCAAGCCACATCTGGAACCCCAAGAGCAGCCCTATCCAG
mouse	CTCATGACCCATATGTGTTTGTAAATGCACCAACACATGCTGGAACCCCAAGAACACCCCTGCAAG
hamster	CTCATGACCCATATGTGTTTGTAAATGCACCAACACATGCTGGAACCCCAAGAACACCCCTATCAAG
mole rat	CTCATGACCCATAGTGTGTTTGTACCGCACCAACACATGCTGGAACCCCGAGAACAACCCCTATAAG
b	
human	L M T H V C F V R H Q P H T G N P K S S P I Q
mouse	L M T H M W F V M H Q P H A G N P K N N P V K
hamster	L M T H M W F V M H Q P H A G N P K N N P I K
mole rat	L M T H R W F V P H Q P H A G N P E N N P I K

Fig. 2 Comparison of the insert exon from various species. **a**, The nucleotide sequence of the IE of the mouse, hamster and blind mole rat aligned with the analogous element from the human αA -crystallin gene. **b**, The deduced amino-acid sequence of the human IE-like element is shown with the IE protein sequences expressed by mouse, hamster and blind mole rat. A deletion in the human DNA sequence is marked by an arrow. Underlined residues in both panels indicate mismatches relative to the mouse sequence.

corresponds to the coding region for the rodent IE, revealing the presence of an IE-like element within the first intron of the human gene.

The nucleotide and deduced amino-acid sequences of the human IE-like sequence were aligned with those of the mouse¹¹, hamster¹² and blind mole rat¹⁵ (Fig. 2). The best alignment, preserving the splice junctions found in the rodent sequences, introduces a deletion in the human sequence (see arrow in Fig. 2). The resulting frameshift brings two stop codons into the reading frame in exon 2 (see underlined codons in Fig. 1). Analysis of the lens proteins of humans and other primates reveals no detectable levels of an αA -crystallin with an insert peptide¹⁰, nor of a truncated αA -crystallin polypeptide (C.J.J., unpublished result; J. Horwitz, personal communication). Apparently, the IE-like element in the human αA -crystallin gene is not a functional protein-coding sequence.

The IE-like element in the human gene is more conserved than the introns but less well conserved than the coding sequences when compared with the rodent genes (see Table 1). The

exon 1 coding sequences in the mouse and human genes are identical in 94% of the nucleotide positions. All but one difference is in the third nucleotide of a codon; only one amino acid out of 63 is different (position 13). In contrast, the human IE-like element is 86% identical in nucleotide sequence to the mouse IE, and has differences nearly equally distributed among the codon positions; for 9 nucleotide changes there are 8 amino-acid differences. The hamster and mouse IE sequences differ by only one amino acid, whereas the blind mole rat, whose eyes are degenerate, has 4 alterations. This suggests that the human IE-like element is not under the same selective pressure as the other exons. By analogy with pseudogenes, we consider the human IE-like sequence to be a silent pseudo-exon.

The presence of the IE-like pseudo-exon in the human αA gene suggests that αA IE arose in early ancestors of many mammals; although useful enough to be retained in some species, it has been silenced in most others. Using the method of Perler *et al.*¹⁶ for calculating diversity between two sequences, and assuming the sequence of the human IE-like element is

Table 1 The divergence of the α A-crystallin genes of the human, hamster and blind mole rat relative to that of the mouse gene

	Exon 1	Intron 1	Insert exon
Human	94	47	86
Hamster	95	59	99
Mole rat	90	49	93

Divergence is shown by comparing nucleotide sequences for the coding region of the first exon, the IE and the intervening sequence between them. Values given for coding regions represent the percentage of positions at which the nucleotides are identical; values for intron sequences represent the percentage of matched residues after alignment by the NUCALN program²² using default values for all parameters.

accumulating changes at the same rate as pseudogenes (4.5×10^{-9} substitutions per site per year)¹⁷, we estimate that the human gene stopped using the IE some 30 to 40 million years ago. Interestingly, the human genome contains pseudogenes for two other lens proteins, both γ -crystallins with functional counterparts in rats¹⁸. This silencing of genes and an exon may have been triggered by similar environmental changes resulting in altered requirements for lens proteins.

The IE was acquired by an extremely well conserved, tissue-specific gene, either by exon shuffling or by the development of functional splicing signals within an intron. Its transformation into a pseudo-exon shows that part of a functional gene can be released from evolutionary pressures while the rest remains stringently conserved. This process may have occurred many

times in the evolutionary history of modern proteins, but is only rarely glimpsed now.

We thank Dr Graeme Wistow, Luke Pallansch and Thomas Lietman for discussion and Dawn Chicchirichi for typing the manuscript.

Received 19 October 1988; accepted 6 January 1989.

1. Gilbert, W. *Nature* **271**, 501–502 (1978).
2. Blake, C. C. F. *Int. Rev. Cytol.* **93**, 149–185 (1985).
3. Holland, S. K. & Blake, C. C. F. *Biosystems* **20**, 181–206 (1987).
4. Breitbart, R. E., Andreadis, A. & Nadal-Ginard, B. *A. Rev. Biochem.* **56**, 467–495 (1987).
5. de Jong, W. W., van Amelsvoort, A. M., van der Ouderaa, F. J. & Bloemendal, H. *Nature new Biol.* **246**, 233–236 (1973).
6. de Jong, W. W., Zweepers, A., Versteeg, M. & Nuy-Terwindt, E. C. *Eur. J. Biochem.* **141**, 131–140 (1984).
7. Cohen, L. H., Westerhuis, L. W., de Jong, W. W. & Bloemendal, H. *Eur. J. Biochem.* **89**, 259–266 (1978).
8. Cohen, L. H., Westerhuis, L. W., Smits, D. P. & Bloemendal, H. *Eur. J. Biochem.* **89**, 251–258 (1978).
9. de Jong, W. W., Cohen, L. H., Leunissen, A. M. & Zweepers, A. *Biochem. biophys. Res. Commun.* **96**, 648–655 (1980).
10. Hendriks, W., Sanders, J., de Leij, L., Ramaekers, F., Bloemendal, H. & de Jong, W. W. *Eur. J. Biochem.* **174**, 133–137 (1988).
11. King, C. R. & Piatigorsky, J. *Cell* **32**, 707–712 (1983).
12. van den Heuvel, R., Hendriks, W., Quax, W. & Bloemendal, H. *J. molec. Biol.* **185**, 273–284 (1985).
13. McDevitt, D. S., Hawkins, J. W., Jaworski, C. J. & Piatigorsky, J. *expl. Eye Res.* **43**, 285–291 (1986).
14. Thompson, M. A., Hawkins, J. W. & Piatigorsky, J. *Gene* **56**, 173–184 (1987).
15. Hendriks, W., Leunissen, J., Nevo, E., Bloemendal, H. & de Jong, W. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5320–5324 (1987).
16. Perler, F. *et al.* *Cell* **20**, 555–566 (1980).
17. Li, W.-H., Gojobori, T. & Nei, M. *Nature* **292**, 237–239 (1981).
18. Meakin, S. O., Du, R. P., Tsui, L. C. & Breitman, M. L. *Molec. cell. Biol.* **7**, 2671–2679 (1987).
19. de Jong, W. W. & Goodman, M. *Sonderdruck Z. Saugetierkunde* **47**, 257–276 (1982).
20. de Jong, W. W., Terwindt, E. C. & Bloemendal, H. *FEBS Lett.* **58**, 310–313 (1975).
21. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
22. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).

Dynamical transition of myoglobin revealed by inelastic neutron scattering

Wolfgang Doster*, Stephen Cusack† & Winfried Petry‡

* Physik Department E13, Technischen Universität München, D-8046 Garching, FRG

† EMBL Grenoble Outstation, Institut Laue-Langevin, 156X, 38042 Grenoble Cedex, France

‡ Institut Laue-Langevin, 156X, 38042 Grenoble Cedex, France

Structural fluctuations in proteins on the picosecond timescale have been studied in considerable detail by theoretical methods such as molecular dynamics simulation^{1,2}, but there exist very few experimental data with which to test the conclusions. We have used the technique of inelastic neutron scattering to investigate atomic motion in hydrated myoglobin over the temperature range 4–350 K and on the molecular dynamics timescale 0.1–100 ps. At temperatures below 180 K myoglobin behaves as a harmonic solid, with essentially only vibrational motion. Above 180 K there is a striking dynamic transition arising from the excitation of non-vibrational motion, which we interpret as corresponding to torsional jumps between states of different energy, with a mean energy asymmetry of 12 kJ mol⁻¹. This extra mobility is reflected in a strong temperature dependence of the mean-square atomic displacements, a phenomenon previously observed specifically for the heme iron by Mössbauer spectroscopy^{3–5}, but on a much slower timescale (10⁻⁷ s). It also correlates with a glass-like transition in the hydration shell of myoglobin⁶ and with the temperature-dependence of ligand-binding rates at the heme iron, as monitored by flash photolysis⁷. In contrast, the crystal structure of myoglobin determined down to 80 K shows no significant structural transition^{8–10}. The dynamical behaviour we find for myoglobin (and other

globular proteins) suggests a coupling of fast local motions to slower collective motions, which is a characteristic feature of other dense glass-forming systems.

Inelastic neutron scattering is a spectroscopic technique which can be used to study protein motions on exactly the same timescale (0.1–100 ps) as is now widely accessible by computer simulation¹¹. Because of the anomalously large incoherent neutron cross-section of the ¹H nucleus, the method specifically probes the motion of hydrogen atoms. As hydrogens are abundant and are uniformly distributed in proteins, the method gives a global view of protein dynamics. The quantity measured experimentally is the incoherent dynamic structure factor $S(q, \omega)$, where $\hbar q$ and $\hbar \omega$ are respectively the momentum and energy transfers between system and incident neutron. $S(q, \omega)$ is the Fourier transform of the time-correlation function of the density fluctuations in a system and can be directly calculated from the results of a molecular dynamics simulation. For samples of myoglobin hydrated with D₂O, we have measured as a function of temperature both the elastic intensity $S(q, \omega \approx 0)$, which gives information on the geometry of motions, and $S(q, \omega > 0)$, which gives the timescale (or spectrum) of the corresponding diffusive and vibrational motion^{12,13}.

Between 4 K and 180 K the elastic intensity of myoglobin has the gaussian form expected for a harmonic solid whose vibrational motion can be described by a Debye–Waller factor, that is, $S(q, \omega \approx 0) = \exp(-q^2 \langle \Delta x^2 \rangle)$, where the mean-square displacement $\langle \Delta x^2 \rangle$ is proportional to temperature (except for quantum effects at very low temperatures). But near 200 K an extra decrease in the elastic intensity at low q is observed, indicating the excitation of new degrees of freedom. The geometry of these motions gives rise to a transition to a non-gaussian elastic intensity (Fig. 1). As the deviation from gaussian behaviour increases with temperature, we can conclude that the motion involves jumping of hydrogens to distinct sites of different energy. The simplest model accounting for these observations involves jumps between two sites separated by a distance d and free energy ΔG (Fig. 2). The powder-averaged normalized elastic

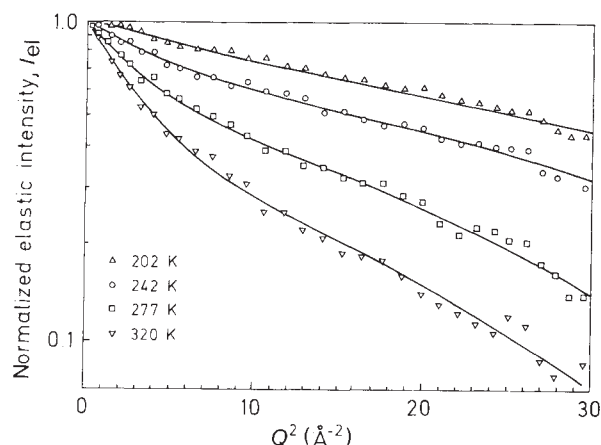


Fig. 1 The elastic intensity $I_{el}(q) = S(q, \omega \approx 0)$ was determined with a fixed window method at a resolution of $10 \mu\text{eV}$ (FWHM) using the backscattering spectrometer²⁶ IN13 at the Institut Laue-Langevin in Grenoble, France with an incident wavelength of $2.23 \text{ \AA}$. The sample was continuously heated from 4 K to 320 K over a period of 24 h. Raw data were corrected for cell scattering and detector response and normalized to unity at $q = 0$. The lines are fits to equation (1). The sample was $\sim 400 \text{ mg}$ sperm-whale myoglobin powder (Sigma) hydrated to $0.38 \text{ g D}_2\text{O}$ per g protein, at which level the protein is essentially fully hydrated. The D_2O contributes $\sim 5\%$ to the total scattering cross-section; the neutron spectrum is thus dominated by scattering from unexchanged protons in the protein (CH , CH_2 and CH_3 and a few inaccessible NH) which reflect internal and surface protein motions.

intensity is then given by¹⁴

$$S(q, 0) = e^{-q^2 \langle \Delta x^2 \rangle_G} \left\{ 1 - 2p_1 p_2 \left(1 - \frac{\sin(qd)}{qd} \right) \right\} \quad (1)$$

The first term describes the gaussian contribution to the mean-square displacement (below 180 K $\langle \Delta x^2 \rangle_G = \langle \Delta x^2 \rangle_{\text{vib}}$) and the term in brackets denotes the elastic incoherent structure factor of the two-state model. The variables p_1 and p_2 denote the probability of finding a hydrogen in the ground or excited state respectively, with $p_2/p_1 \propto \exp(-\Delta G/RT)$. The second factor contributes to the q -dependence of $S(q, 0)$ essentially only for $qd < 1$. The hydrogen mean-square displacement is given by

$$\langle \Delta x^2 \rangle = - \left(\frac{d \ln \{ S(q, 0) \}}{d(q^2)} \right)_{q=0} = \langle \Delta x^2 \rangle_G + \frac{p_1 p_2 d^2}{3} \quad (2)$$

which is the initial slope of the curves in Fig. 1. Least-squares fits of equation (1) to the data (Fig. 1) give $p_1 p_2$, $\langle \Delta x^2 \rangle$ and $\langle \Delta x^2 \rangle_G$ as a function of temperature (Fig. 2). A van't Hoff plot of p_2/p_1 gives an energy asymmetry of $\Delta H = 12(\pm 2) \text{ kJ mol}^{-1}$ and entropy $\Delta S/R = 3.0(\pm 0.1)$. The value of the jump distance d is found to be $1.5(\pm 0.1) \text{ \AA}$. The magnitude of this value would indicate the involvement of torsional degrees of freedom and large amplitude, fast dihedral angle fluctuations are indeed observed in molecular dynamics simulations of proteins¹⁵. As discussed more fully elsewhere, the parameters given above are dependent on hydration. In particular, the excess entropy of the excited state is significantly reduced on dehydration and the distance d drops to $1.1 \text{ \AA}$. Furthermore at low and intermediate levels of hydration, $\langle \Delta x^2 \rangle_G$ is dominated by the vibrational contribution $\langle \Delta x^2 \rangle_{\text{vib}}$, whereas in the high-hydration data presented here there is an additional gaussian process, $\langle \Delta x^2 \rangle_G = \langle \Delta x^2 \rangle_{\text{vib}} + \langle \Delta x^2 \rangle_\alpha$. The resulting $\langle \Delta x^2 \rangle_\alpha$ is also shown in Fig. 2.

The extra loss in elastic intensity is compensated by neutrons scattered into a broad quasielastic spectrum between 0 and 3 meV . To derive the actual shape of this spectrum requires a careful subtraction of the vibrational background (manuscript

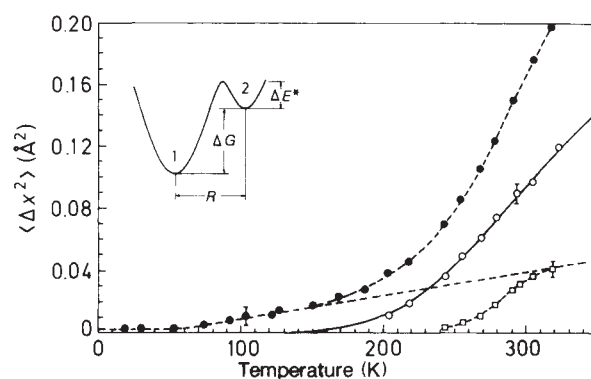


Fig. 2 Average mean-square hydrogen displacements obtained from the analysis of $I_{el}(q)$ assuming an asymmetric two-state model. The full circles give the total mean-square displacement $\langle \Delta x^2 \rangle$ defined by equation (2). The solid line shows $\langle \Delta x^2 \rangle_\beta = p_1 p_2 d^2/3$ obtained using the parameters in the text and the open circles give the corresponding experimental points. The squares show $\langle \Delta x^2 \rangle_\alpha$ (see text). The dashed straight line is the vibrational contribution $\langle \Delta x^2 \rangle_{\text{vib}}$ extrapolated linearly from low temperature. A symmetrical model with two excited states may be more appropriate for torsional motion. Such a model does indeed fit the data equally well, but the derived parameters (apart from the entropy, which is lower) are not significantly different. Jump models with multiple ground states are excluded by the data. Quantum effects are important below 50 K , leading to a temperature-independent $\langle \Delta x^2 \rangle$ of $0.005 (\pm 0.005) \text{ \AA}^2$.

in preparation; see also ref. 16). In Fig. 3 we show the resultant quasielastic spectra combining data from two spectrometers with different energy resolutions and the inset shows the corresponding correlation functions. Two spectral components with different shape and temperature dependence can clearly be recognized, a fast β -process and a slower α -process. The increase with temperature of the intensity of the broad β -line is consistent with local jumps between two states of energy asymmetry of 12 kJ mol^{-1} . Surprisingly however, the linewidth is temperature-independent with a value of $1.5\text{--}2 \text{ meV}$ (corresponding to a correlation time of $\sim 0.3\text{--}0.5 \text{ ps}$, see inset to Fig. 3). Within the framework of the two-state model, this would imply that the activation energy ΔE^* , essentially the depth of the well in the excited state, is smaller than RT ($< 1.6 \text{ kJ mol}^{-1}$). Qualitatively similar behaviour of the quasielastic spectrum has been obtained with the polymeric glass polybutadiene¹⁷.

The linewidth of the α -relaxation by contrast broadens with increasing temperature. Below 240 K this component is not well resolved and therefore contributes to the elastic intensity. The gaussian mean-square displacement $\langle \Delta x^2 \rangle_\alpha$ (Fig. 2) is related to this process. The shape of the α -relaxation is clearly not lorentzian but is reasonably well fitted by a Cole-Davidson function¹⁸ often used to parameterize the α -process in liquids and polymers:

$$S(q, \omega) = -A_1(q) \text{Im} \{ (1 + i\omega\tau_0)^{-b} \} / \omega \quad (3)$$

This formula implies a power-law dependence at high frequencies with exponent $-(1+b)$, (a lorentzian has $b=1$). We find a value for b of $0.25(\pm 0.1)$, which is temperature independent between 250 K and 350 K (Fig. 3). The temperature-independent shape essentially excludes an explanation of this component in terms of a distribution of correlation times unless we assume that the corresponding distribution of activation energies varies with temperature in a unique way. The shape of the total spectrum (α and β) however is qualitatively consistent with the results derived from a mode-coupling theory of the liquid-glass transition in a hard core liquid¹⁹⁻²². The α -process in this theory is a collective effect, its shape arising as a result of non-linear coupling between density fluctuations. The β -process corresponds to local motions of atoms in the cage formed by their neighbours. Our data therefore demonstrate that diffusive motions in a protein resemble those in other close-packed sys-

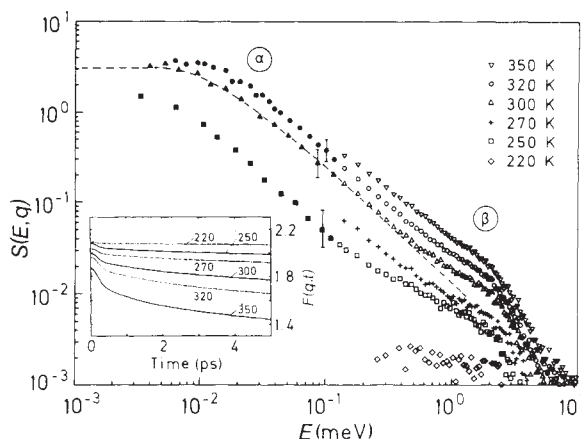


Fig. 3 Log/log plot of quasielastic neutron spectra obtained by combining the results of IN13 (filled symbols) and the time-of-flight spectrometer IN6 (open symbols)²⁶. IN6 has a gaussian resolution function ($\sigma = 50 \mu\text{eV}$) and an incident wavelength of $5.1 \text{ \AA}$. In addition to the usual corrections, a vibrational background has been subtracted at each temperature; this was obtained by scaling the inelastic scattering measured at 180 K or 100 K , assuming harmonic behaviour. The IN6 data represent a single scattering angle (58.5°); the IN13 data have been averaged over the q -range of IN6 ($2 \text{ \AA}^{-1}$) to improve the statistics. Shape and linewidth of the spectra are roughly independent of q . The data from the spectrometers were matched at the overlap energy of 0.1 meV . Spectra are not corrected for multiple scattering. The effect of instrumental resolution is small within the energy range presented here. The dashed line represents equation (3) convoluted with the resolution function of IN13, with $b = 0.25$ and $\langle \tau \rangle = b\tau_0 = 2.10 \times 10^{-11} \text{ s}$. The β -process has a shape intermediate between gaussian and lorentzian and an estimated correlation time of 0.35 ps . Inset: correlation functions obtained from the IN6 quasielastic spectra by Fourier transform and deconvolution of the instrumental resolution function. The overall decrease in intensity with increasing temperature is due to the decrease in the elastic peak which was not subtracted out and contributes a flat background to the correlation functions.

tems dominated by hard-core interactions. But some unique features have to be explained in terms of protein structure: the exponent 0.25 is unusually small (close to $1/f$ noise) and the non-gaussian scattering law, indicating jumps between distinct sites of different energy, suggests that the geometry of the local motion is highly constrained by strong interactions along the chain. A similar asymmetric two-state model has been proposed to explain the microwave absorption of the main-chain in lysozyme and haemoglobin²³. The corresponding energy asymmetry is almost identical to our result. In this connection we remark that neutron data on hydrated lysozyme powders (not shown) behave very similarly to those from myoglobin, suggesting that the dynamic phenomena described here are a general property of globular proteins. The temperature dependence of the hydrogen mean-square displacement (Fig. 2) resembles that derived for the heme iron by Mössbauer spectroscopy³⁻⁵. This correlation indicates that the picosecond fluctuations observed with neutron scattering could couple to the heme group and are likely to be an important component in the decrease of the Lamb-Mössbauer factor above 200 K (ref. 5), as well as in the slower motions resolved in the Mössbauer spectrum (down to 10^{-7} s). Further discussion of this point awaits the current re-evaluation of new and more accurate Mössbauer data²⁴.

Finally we suggest that an important test of the molecular dynamics simulation technique would be to reproduce the temperature dependence of the picosecond protein dynamics reported here. This would also provide insight into the specific atomic motions involved. A step in this direction is the recent theoretical analysis of multiple conformational states in myoglobin²⁵.

This project is supported by a EMBO short-term fellowship and a grant from the Deutsche Forschungsgemeinschaft. We

thank Dr A. Dianoux for assistance on IN6 and discussions on the manuscript. Assistance and useful suggestions by Dr Fujara, Professor Dr E. Lüscher and F. Post are also gratefully acknowledged.

Received 28 September 1988; accepted 9 January 1989.

1. McCammon, J. A. *Rep. Prog. Phys.* **47**, 1-46 (1984).
2. Levy, R. M. *et al. Biophys. J.* **48**, 509-518 (1985).
3. Parak, F. *et al. FEBS Lett.* **117**, 368-372 (1980).
4. Keller, H. & Debrunner, P. *Phys. Rev. Lett.* **45**, 68-71 (1980).
5. Bauminger, E. R., Cohen, S. G., Nowik, I., Ofer, S. & Yariv, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 736-740 (1983).
6. Doster, W., Bachleitner, A., Dunau, R., Hieble, M. & Lüscher, E. *Biophys. J.* **50**, 213-219 (1986).
7. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsales, I. C. *Biochemistry* **14**, 5355-5373 (1975).
8. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. *Nature* **280**, 558-563 (1979).
9. Hartmann, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4967-4971 (1982).
10. Parak, F. *et al. Eur. Biophys. J.* **15**, 237-249 (1987).
11. Cusack, S. in *The Enzyme Catalysis Process: Energetics, Mechanism and Dynamics* (eds Cooper, A. & Houben, J. (NATO ASI series, Plenum, London, in the press).
12. Springer, T. in *Quasielastic Neutron Scattering for the Investigation of Diffusive Motions in Solids and Liquids* (Springer Tracts in Modern Physics, Vol. 64, 1972).
13. Bee, M. in *Quasielastic Neutron Scattering: Principles and Applications in Solid-state Chemistry, Biology and Materials Science* (Hilger, Bristol, 1988).
14. Stöckli, H., Furrer, A., Schoenberg, Ch., Meier, B. H., Ernst, R. R. & Anderson, I. *Physica* **136B**, 161-164 (1986).
15. Levitt, M. *J. molec. Biol.* **168**, 621-657 (1983).
16. Cusack, S., Smith, J., Finney, J., Tidor, B. & Karplus, M. *J. molec. Biol.* **202**, 903-908 (1988).
17. Frick, B., Richter, U., Petry, W. & Buchenau, U. *Z. Phys.* **B70**, 1-3 (1988).
18. Davidson, D. W. & Cole, R. H. *J. chem. Phys.* **18**, 1417-1419 (1950).
19. Bengtzelius, W., Götze, W. & Sjölander, A. *J. Phys.* **C17**, 5915-5933 (1984).
20. Götze, W. & Sjölander, L. *J. Phys.* **C20**, 879-894 (1987).
21. Götze, W. & Sjölander, L. *J. Phys.* **C21**, 3407-3471 (1988).
22. Doster, W., Cusack, S. & Petry, W. *Proc. ILL Workshop Dynamics of Disordered Materials*, Grenoble, 1988.
23. Poglitsch, H., Kremer, F. & Genzel, L. *J. molec. Biol.* **173**, 137-142 (1984).
24. Parak, F., Heidemeier, J. & Nienhaus, G. U. *Hyperfine Interactions* **40**, 147-158 (1988).
25. Elber, R. & Karplus, M. *Science* **235**, 318-321 (1987).
26. *Neutron Research Facilities at the ILL High-Flux Reactor* (Institut Laue-Langevin, Grenoble, 1988).

GUIDE TO AUTHORS

GENERAL

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully. Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50-80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963-65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.